



**CENTRE DE RECHERCHE
SUR LES TRANSPORTS
(C.R.T.)**

*CENTRE FOR RESEARCH
ON TRANSPORTATION*

UNIVERSITÉ DE MONTRÉAL
ÉCOLE DES HAUTES ÉTUDES COMMERCIALES
ÉCOLE POLYTECHNIQUE



CENTRE DE RECHERCHE SUR LES TRANSPORTS
UNIVERSITÉ DE MONTRÉAL
C.P. 6128, SUCCURSALE A,
MONTRÉAL
CANADA H3C 3J7

TÉLÉPHONE: (514) 343-7575
TÉLEX: (21) 05-561325
EMAIL: crt@crt.umontreal.ca
TÉLÉCOPIEUR (FAX): (514) 343-7121

A paraître dans Operations Research.

CRT-713

**A Dynamic Stochastic Model for the
Allocation of Empty Containers**

by

Teodor Gabriel Crainic^{1,2}

Michel Gendreau¹

Pierre Dejax³

¹ Centre de recherche sur les transports, Université de Montréal, C.P. 6128, succursale A, Montréal, Canada H3C 3J7.

² Département des sciences administratives, Université du Québec à Montréal, C.P. 6192, succursale A, Montréal, Québec H3C 4R2

³ Laboratoire Economique, Industriel et Social, Ecole Centrale Paris, Chatenay-Malabry, France.

Dépôt légal – 4ème trimestre 1991

ABSTRACT

The empty container allocation problem occurs in the context of the management of the land distribution and transportation operations of international maritime shipping companies. It involves dispatching empty containers in response to requests by export customers and repositioning other containers to storage depots or ports, in anticipation of future demands. We propose for this problem a dynamic, stochastic model which accounts for its specific characteristics: the space and time dependency of events and the uncertainty of supply and demand data, substitutions, relationships with partner companies, imports and exports, massive equilibration flows, etc. We give a mathematical formulation and show an equivalent large stochastic multicommodity network flow with gains model. Various modeling choices, data requirements, and algorithmic considerations related to the implementation of the model are also discussed.

KEYWORDS Dynamic, Stochastic Models;
 Multicommodity Network Models with Gains;
 Vehicle Allocation; Fleet Management;
 Land Container Transportation.

RESUME

Le problème d'affectation de conteneurs vides se pose dans le contexte de la gestion des opérations de distribution et de transport terrestre des compagnies maritimes internationales. Il s'agit d'envoyer des conteneurs vides aux clients exportateurs qui en font la demande ainsi que d'en repositionner d'autres, à des dépôts intérieurs ou portuaires, afin d'être en mesure de satisfaire les demandes futures. Nous présentons pour ce problème un modèle dynamique et stochastique qui reflète ses caractéristiques propres: les dépendances spatio-temporelles des offres et des demandes, les substitutions, les échanges avec d'autres compagnies, les importations et exportations, l'existence de flots massifs d'équilibrage inter-dépôts, etc. Nous présentons une formulation mathématique ainsi que le modèle équivalent de multiflots stochastiques avec gains. Nous discutons également de certains choix de modélisation, des exigences de disponibilité de données, et des perspectives algorithmiques, en vue d'applications pratiques du modèle.

MOTS CLES Modèles dynamiques et stochastiques;
 Modèles de flots dans les réseaux avec gains;
 Affectation de véhicules;
 Gestion de flotte; Transport terrestre de conteneurs.



INTRODUCTION

Since the seventies, freight transportation activities have seen many fundamental changes both in terms of technology and of organization. In particular, one must note the increasing use of multimodal transportation routes and of containerized transportation practices. This is particularly true in the area of international maritime shipping, where the maritime carrier normally owns the containers and insures their multimodal transportation from the initial origin to their final consignee.

We are particularly interested in the planning of the land transportation part of the international maritime shipping of containers. This is an extremely complex activity, especially if one aims to simultaneously optimize the cost and service aspects of the company's operations in a competitive environment. A fundamental component of this problem is concerned with the allocation of empty containers to customers, in preparation of future loaded export movements, and the repositioning of other empty containers to storage depots or ports to be used to fulfill next periods' orders (both expected and forecast).

The goal of this paper is to propose a dynamic, stochastic model for empty container allocation in a land distribution and transportation system, which fully responds to the complexity of the specific problems which arise in this area, particularly the space and time dependency of events and the uncertainty of supply and demand data, plus specific operational characteristics such as substitutions, relationships with partner companies, imports and exports, massive equilibration flows, etc. The focus of the paper is on modeling issues, our aim being to present a formulation as general as possible for this class of problems.

To our knowledge, there does not exist any other model specifically aimed at the empty container allocation problem that we study. In a recent survey of the literature on fleet management models in freight transportation, Dejax and Crainic (1987) noted that relatively little effort was directed toward developing models aimed specifically toward container transportation issues, with the bulk of the work in the area being dedicated to the maritime aspects of the problem. Thus, Dejax and Crainic quote a few authors who addressed the problem of the empty container distribution as that of allocating containers available in surplus at a terminal, from earlier loaded shipments, to demanding terminals (e.g. a customer's warehouse, a rail yard, or a port), in preparation for subsequent loaded movements.

Anthonisse (1988) is a recent example of such efforts, for the worldwide maritime container fleet management problem.

The state of the art described by Dejax and Crainic may be explained, at least partially, by the combined effect of the relative complexity of the container land transportation issues and the opinion, prevalent until very recently in the industry, that the impact of land operations on the overall company's performance was marginal. This perception is, however, gradually evolving and, especially in today's increasingly competitive environment where growing customer demand is associated to the search for cost-efficient high-quality-service operations, the land portion of the container journey is receiving increasing attention. This view is similar to those of, among others, Jarke (1982), who reviews container management problems and who insists on the necessity for efficient information and decision-support systems for container transportation logistics, and Antush (1989), who emphasizes the need for developing new information systems and planning methodologies for container transportation.

The paper is organized as follows. In the first section, we give a brief description of the issues related to planning container distribution and transportation systems, as well as of the general methodological framework we propose to address them. The land container allocation problem, its main components and characteristics, is described in Section 2. Section 3 is dedicated to the development of the mathematical and network formulations of the deterministic, single commodity version of the model, while the stochastic version of this model is presented in Section 4. Substitutions among container types are treated in Section 5 which thus contains the more general, multicommodity, formulations of the previous models. In Section 6 we compare our modeling framework with the work of other authors, while Section 7 is dedicated to considerations related to the practical implementation of the model. Our conclusions indicate future research directions in this area.

1. PROBLEM CONTEXT

Our research is motivated by the analysis of typical land transportation problems encountered by international shipping companies that operate maritime and land networks of considerable, even continental, scale. Crainic, Gendreau and Dejax (1990) describe the various operations and planning issues involved, and present the methodological framework that we propose in order to address them. To facilitate the understanding of the allocation model, this section is dedicated to a brief overview of these issues and design principles.

Arriving ships carry containers, which generally come in two sizes and about ten major types each, loaded with imported goods, as well as empty containers returning from previous exports. Loaded containers must further be transported to their final destination, usually an industrial firm or a distribution center located somewhere in the continent. Ground transportation may be performed by using rail, truck, barges and mixed modes. This latter option usually consists of an initial railway shipment to an inland railway terminal, followed by a final movement by truck to the customer. The empty containers that are imported may either stay at the port for a while, or be immediately dispatched wherever they are needed for subsequent exports.

Once unloaded, empty containers have to be picked up at the customer's site and moved away. They may either return to the port of origin, or may be transported to any other depot: port or inland terminal. This return shipment may, but more often than otherwise will not, be made using the same vehicle (truck, railway car, ...) that has been used for the transportation from the port to the customer. Similarly, exporting firms need empty containers to be delivered to them, either from the future export port, or from any other location. Due to complex organizational considerations, very few (Dejax et al., 1988, report some 3% for the case they studied) empty container movements are made directly from import to export customers. After loading, containers are transported to the export port, and are loaded on ships together with empty containers sent abroad to cope with the worldwide imbalance in the supply/demand of certain container types.

Note the importance of empty container movements. First, every commercial (profitable) movement of a loaded container generates almost automatically a non-profitable empty container movement: once loaded containers are delivered to a customer's site and are unloaded,

they normally have to be moved away empty, while empty containers have to be delivered to exporting customers. Moreover, significant regional imbalances between imports and exports within the area served by the company (their amplitude may vary according to the container type and the geographical subdivisions considered), as well as opportunities for cost reductions due to mass transportation of containers over relatively long distances (by railway, for example), result in land movements of empty containers directly between depots: these are called *balancing flows*.

Thus, by their impact on the amplitude of the resource requirements and on the quality of service offered to the company's customers, empty movements play a central role in the management of the land distribution and transportation of containers. A few figures to illustrate and support this point (Dejax et al., 1988). A major European container shipping company operates worldwide maritime shipping lines to and from some twenty European ports, while its land distribution and transportation network mainly covers France, Germany, Holland, Belgium and Luxembourg. In 1986, its operations involved over 300,000 land container movements, which represented an estimated total distribution and transportation cost of some U.S. \$ 50,000,000. Of these, 40% were empty movements!

Several key questions have to be addressed by container shipping companies for the organization of their land distribution and transportation operations, questions which clearly belong to several planning levels: strategic (e.g. the selection of inland terminals and the determination of their logistic zones), tactical (e.g. the evaluation of interdepot empty flows) and operational (e.g. the dynamic dispatching of containers). They cannot all be represented by a single model. Moreover, they do not require the same data, nor the same personnel and resources for their planning and execution. Hierarchical relations exist among them and are reflected in the flow of information and decisions. Yet, the modeling effort is significantly facilitated by the fact that container shipping companies do not often need to build expensive inland terminals: they may in fact use already existing terminals, such as rail yards, for a cost much lower than the cost of building their own facilities. This allows them to periodically re-evaluate the decisions concerning the use of inland terminals and the corresponding allocation of customers to depots. An integrated, multilevel methodology may then be designed in order to attempt to answer these questions. We have developed such a methodology, which reflects the observed hierarchy in the decision process and the flow of information.

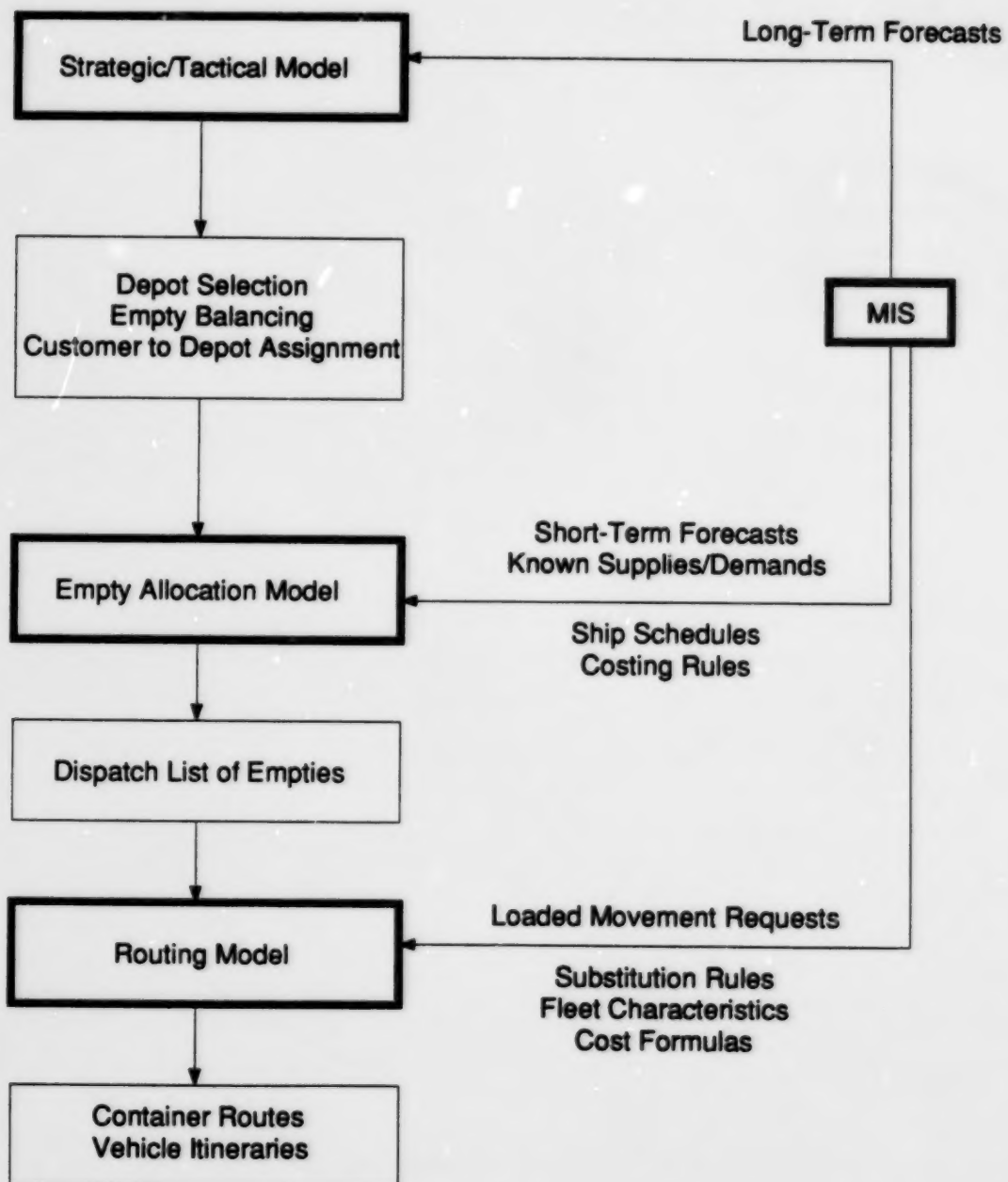


Figure 1. Outline of the Overall Planning Approach

Figure 1 displays a schematic representation of this approach. The topmost level is concerned with the strategic/tactical planning of operations. The model is formulated as a multimode, multicommodity location/distribution problem with interdepot balancing requirements (Crainic, Dejax and Delorme, 1989). Several specifically tailored algorithms (see Crainic, Gendreau and Dejax, 1990, for references) have been developed, which efficiently make full use of the particular characteristics of the formulation. The main output of this phase is the *strategic/tactical transportation plan*, which specifies the network configuration for the current planning period: the depots that are to be used, the allocation of customers and customer zones to depots, and the main interdepot balancing flows. Balancing traffic is indicated by container type, while the allocation of customers to depots is also specified by container type and, additionally, by direction of movement, from customer to depot and from depot to customer. In general, a customer may be assigned to several depots, for each type of container and each direction of movement. Then, the strategic/tactical plan indicates the relative importance (proportions) of these flows.

The two other models correspond to the level of the operational planning of the company's activities. At this level, one has to make sure that demand is satisfied and that the most effective routes and means of transportation are selected and used. Major factors that characterize this level of planning are the dynamic and stochastic environment and the guidelines established by the strategic/tactical plan; one generally aims towards operational decisions, taken locally in a dynamic environment, which achieve the best compromise between short term and global, long term optimization. Ideally, one would like to develop a single mathematical model to optimize the short-term land operations of the company, in order to fully account for the interactions between the various decisions to be made. Given the intrinsic complexity of the problem at hand and the current state of the art in O.R., this is not feasible. This explains our decision to break down this large problem, and to develop two separate optimization models.

The allocation model aims at the determination of the "best" distribution of empty containers that satisfies both known and forecast customer demands. Economic criteria guide the search for solutions which have to respect the strategic/ tactical objectives and the policies of the company. To allow for the high level of uncertainty inherent in the real-time operation of a transportation system and for the impact of current decisions in the future, it is designed as

a stochastic, dynamic model which covers a planning horizon of one or two weeks. The objective of this paper is to present the formulation of the allocation model.

The routing model strives to minimize the overall transportation cost of the loaded and empty containers from their origin to their destination. This is a complex and difficult problem, since it involves not only the determination of the "best" (multimodal) container itineraries, but also the construction of the routes for the vehicles required to transport them (for one or several modes). It is worthy to note that, since the container shipping company does not own, in any significant measure, the required land transportation means, substantial savings can be achieved by using a routing model. This is particularly true for truck transportation where most savings are achieved by generating multi-stop "circuit" routes (and by suitably matching requested container movements to these routes) for which tariff reductions may be negotiated. The output of the routing model is a list of movement orders which completely describe the loaded and empty movements to be executed during the next period (Crainic et al., 1991).

As illustrated in Figure 1, the use of these models requires that a wide gamut of information be retrieved from the company's information system. In particular, data on demand is to be given, including long and short-term forecasts and known supplies, and demands for the forthcoming periods. Equally essential are the rules that specify how transportation means are chosen and priced, plus the company's own accounting criteria concerning the various costs related to owning and using containers. In the following sections, we further discuss these information needs in relation to the requirements of the allocation model.

2. PROBLEM DESCRIPTION

Several decisions are interwoven into the empty container allocation process. These decisions are mainly concerned with *when* to initiate a transportation operation, as well as with the *from where* to originate, or *where* to terminate, the operation. (At this planning stage, the *how* to perform the movement – mode, route, etc. – is usually implicit in the spatial and temporal definition of the movement.) Besides these “movement” decisions, one is also concerned with operations that affect the availability of containers for allocation: stock levels at depots, substitutions among containers of various types, additional resources (e.g. new or borrowed containers), etc. These decisions require a great deal of information, including a variety of considerations regarding the company's distribution and transportation system, the levels of supplies and demands, the company's operating policies and relations with its environment (competition, partner shipping companies, official rules and regulations, etc.). Here, we describe the main decisions, system components and data considerations which make up the empty container allocation process. The emphasis is on the elements which most strongly characterize (and increase complexity of) the problem, and hence which have significantly influenced the design of the model.

The *physical* network which corresponds to the company's distribution and transportation system is made up of ports, depots and customers, as well as of the transportation links connecting these elements. Such a network is very schematically drawn in Figure 2.

In a land container distribution and transportation system, *ports* play a special role. They are the main entry and exit points of the system, since most of the exchanges between the land transportation system and the international shipping network which contains it take place in ports. Furthermore, container shipping companies are primarily maritime industries; hence, they operate significant parts of their land network from ports, where they also maintain important stocks of empty containers. Consequently, we have decided to specifically include ports in the formulation.

“Inland” *depots* complete the backbone of the land distribution system. Physically, these depots often are located in facilities offered by other transportation modes (e.g. rail yards), the container shipping company only renting the space it needs. Consequently, the company may adjust its network configuration and operations to the demand variations. Yet,

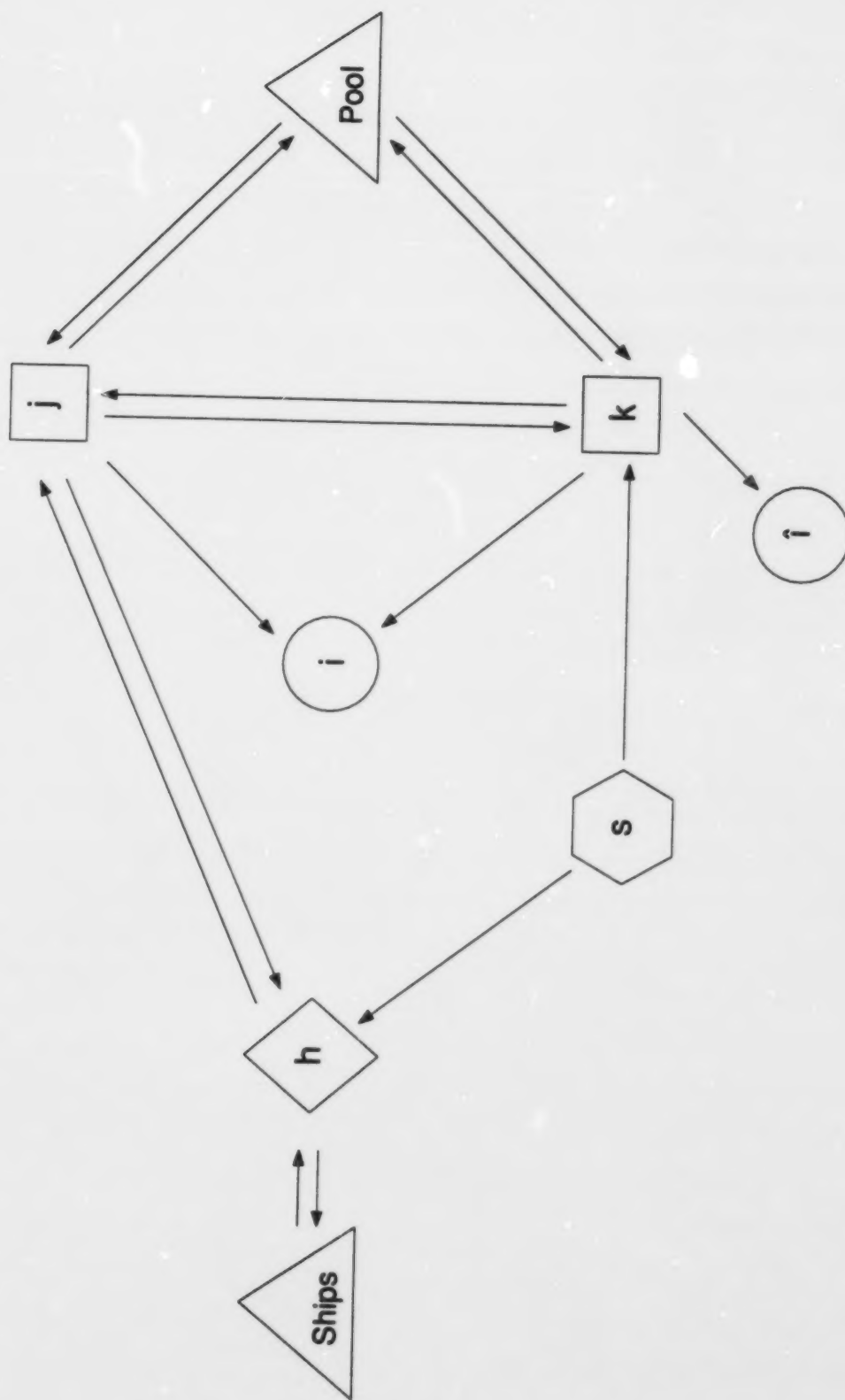


Figure 2. Possible Container Movements in a Land Transportation System

for a given planning horizon, the port and depot network configuration (depots j , k and port h of Figure 2) is fixed.

Depots, both inland and in ports, serve as primary storage and transit locations for empty containers (they may also serve as intermodal transfer points for multimodal transportation of loaded containers; this aspect is, however, quite beyond the scope of this paper): stocks of empty containers are built, demand's for empties are satisfied by allocating containers out of these facilities, empty containers available at customers' sites are shipped to depots, balancing movements are initiated from there. The warehousing function and decisions are rather straightforward. In order to describe the movement decisions, however, we need first to introduce the *customers* of the system.

We distinguish between *demand* and *supply* customers (although some may be both). A demand customer requires that empty containers be delivered at its site, while a supply customer has empty containers, from a previous loaded import shipment, that have to be taken away. The number of containers, their type, as well as temporal conditions (e.g., the latest date of delivery) characterize each demand and supply request.

The location of a customer's site is known. For each demand customer, the strategic/tactical plan had also determined the depots that may satisfy its requests for each type of container. In Figure 2, for example, demand customer i may be served from both depots j and k , while the requests of customer $\hat{i}$ may only be satisfied out of depot k . For each customer demand, one also generally knows the *delivery window* within which containers have to be delivered. The delivery windows are mainly determined by the due date in ports (determined by the ship schedules) and by estimations of the time necessary to load the containers and to transport them. It may be worth noting that, at the level of planning considered herein, the definition of delivery windows implies that backorders are not allowed in the system. The allocation decision hence corresponds to determining how to satisfy the customer demands: what quantities, from what depot and when to initiate the movement.

Supply customers may either send the empty containers to one (or several) of their assigned depot(s) (depot k for customer s of Figure 2), or back to the port that originated the loaded movement (link (s, h) of Figure 2). This is an important issue, due to its influence on the inventory levels at depots and on the subsequent allocation decisions. Thus, the input of

loaded containers into the system and the moment when these containers are again available to the company have a significant impact both on routing issues and on the future *supply* of empty containers at depots.

The *dynamic* and *stochastic* characteristics of the problem are mostly evident in the definition of the supplies and demands of the system. The *empty container supply* consists of all the containers which are made available for allocation at any given time. The supply for each type of container is made up of the stocks at depots, of the containers present at supply customer sites, and of those entering the system via arriving ships in ports. Note, however, that before being available for allocation, most containers have to be moved: from supply customers to some depot, or between depots. Furthermore, not all customer supplies become available at the same moment, even if the loaded import containers are delivered from the same ship. This is mainly due to uneven transportation and unloading times.

The *empty container demand* is indicated for each customer, for each type of container, and for each period of the planning horizon. In fact, the same customer may require containers, of the same or different types, at several periods. Therefore, one needs to distinguish between the "physical" customer (its site) and its various demand requests. Most of these demands correspond to *identified* customers, with known sites, and can be considered explicitly in the allocation model. However, while some of these demands are completely determined ("sure" orders), those of some other customers are known with less certainty, especially as the period where the demand is expected is farther down the temporal axis. Finally, it is also known that, in each region and for each type of container, there is a nonzero probability that an unforecast demand will pop up in some future period. We associate this type of demand with the depots.

As previously described, the *strategic/tactical plan* specifies the network configuration for the current planning period, and the main pairs of depots between which interdepot empty container flows are required. This strategic/tactical plan represents the set of operational policies in an ideal situation. In practice, however, one has to deal with the uncertainties and the stochasticity inherent to real operations, and has therefore to depart from the target represented by the plan. In particular, movements of individual containers may take place between depots which are "close by", and balancing flows have to be adjusted accordingly. Hence, when exactly to initiate the interdepot balancing flows, as well as their amplitude, is a

decision that belongs to the scope of the allocation model.

Differences between supplies and demands of containers of given types exist not only at the national and regional levels, but also at the level of the international trade. Therefore, one notes the existence of *external movements of empties*, associated with the input/output function of ports and represented by the arrival (departure) of empty containers from (to) outside the land system. Ships perform these movements, together with the import and export of loaded containers. In the process we are studying, one has no (for all practical purposes) possibility to modify the ships' schedules when planning the land distribution and transportation system. Hence, these schedules are assumed known, as are the manifests of the ships specifying the origin, destination, type and number of each shipment of (loaded or empty) containers. Consequently, in our formulation, ship arrivals and departures correspond to temporal entry and exit points of the system, determining when certain supplies become available and the latest date when some demands may be satisfied. We do not track containers outside the land transportation system.

Consideration has also to be given to the *external affairs* activities which impact upon the availability of empty containers, such as possible exchanges with partner container shipping companies, renting of empty containers, acquisition of new ones, etc. The wreckage of containers, either definitely (when they are scrapped and go out of the system) or temporarily (an estimation of the expected repair time is then needed), falls also into this category. We schematically illustrated these possibilities by the connections between depots j and k (the link from port h is omitted for clarity's sake) and the node *pool* of the network in Figure 2. Probability distributions based on historical data may be used to represent the amplitude of these activities.

An important characteristic of the problem, and one which adds considerably to its degree of complexity, is the possibility to *substitute* one type of container for another. The container *substitution rules* are specific for each pair of container types, and are not necessarily symmetric, neither in type (type A may be substituted for type B, but the reverse is not permitted), nor in number (one type A container may be substituted for a container of type B, but it takes two of these to replace one type A). Moreover, these rules may have to be adjusted according to the loading/unloading facilities available at the customer's site (one cannot substitute a 40-foot container for two 20-foot ones if the dock is not long enough), and

the specific commodity that is to be transported (one cannot substitute two 20-foot containers for a 40-foot one if anything lengthier than twenty feet has to be transported). The handling of substitutions is further complicated by the emergence of containers of dimensions which differ from the "traditional" forty or twenty feet.

To conclude this problem description, we summarize the dynamic aspects that characterize the empty container allocation problem: the temporal characteristics of the problem (of various transportation activities, modes and itineraries, port and terminal operations and delays, etc.), the need to reposition (and substitute) containers in order to be able to satisfy future demands, the fact that demands and supplies "appear" at various points in time and vary with time (even for the same "physical" customer), the existence of delivery windows for demand customers, and the stocks of containers at depots used as buffers in the face of varying supplies and demands. Concerning the stochastic aspect of the problem, one notes that it derives from the uncertainties of future supplies and demands. Hence, even though only immediate movements and decisions are of actual interest and are transmitted to the routing module, a much longer planning horizon is required in order to properly represent and assess their impact. Finally, note that the substitution mechanism, the stocks at depots, and the possibility to access containers outside the system (buy, rent, etc.), forms a very rich recourse structure for the problem.

The structure of the formulation that we propose in order to solve this problem may be described as a *stochastic dynamic network* model over a *rolling horizon* of n days or periods. This corresponds to a multicommodity space-time description of possible empty container movements and operations (holding at nodes, substitution decisions, etc.), each container type corresponding to one commodity. To make for a clearer and smoother presentation, we introduce the formulation in three stages. First a deterministic model of the single commodity case, followed by its stochastic formulation. Finally, we introduce the general problem with multicommodities and substitutions. For all modeling variants, we start with the formal presentation of the mathematical formulation, followed by the description of the corresponding network model. All formulations are dynamic, so this characteristic is not included into the section titles.

3. DETERMINISTIC SINGLE COMMODITY MODEL

This section is dedicated to the presentation of the modeling of the problem as a deterministic, dynamic, single commodity formulation. It is an idealized situation which allows nevertheless a good presentation of the notation and mathematical formulation. Most concepts and notation naturally generalize to the other two models. Note that if this formulation is to be used *per se*, a number of additional questions should be addressed, including the treatment of the end of the horizon. We discuss this issue in Section 7.

3.1 The Mathematical Model

For a given planning horizon, the configuration of the port and depot network which is to be used by the company is invariant with the specific period (day). On the other hand, the set of "active" customers (i.e. customers with outstanding requests for empties or with available empty containers) does vary according to the time period. Furthermore, requests by the same firm with different due dates have to be treated separately. The set of *identified customer requests* is thus specifically defined for each time period, according to the strategic/tactical plan and the forecast demand for known customer firms. Note that these definitions imply that two requests of the same customer are treated as two different items by the system. Let

- T : Length of the planning horizon, $t = 1, 2, \dots, T$.
- H : Set of ports (harbors) – entry/exit points of the system.
- D : Set of "inland" depots (does not include the port depots).
- I^t : Set of *identified demand customer requests*, to be delivered at the latest in period t , $t = 1, 2, \dots, T$.
- Δ_i : *Delivery window* of customer i , $i \in I^t$:
 if $\Delta_i = 0$, one has to deliver in period t exclusively;
 if $\Delta_i > 0$, one has the option of delivering in all periods $t, t - 1, \dots, t - \Delta_i$.
- S^t : Set of *supply customer requests* which ask in period t , $t = 1, 2, \dots, T$, that the empty containers present at the customers' sites be picked up.

Note that we assign ports and non-port ("inland") depots to different sets; this makes for a cleaner notation, especially for latter versions of the model. For plain descriptions however, we shall use the term "depot" to indicate any facility, port or inland depot, where the

warehousing function is performed.

The network operations are driven by the amount of empty container supply and demand associated with customers, depots and ports, for each one of the periods of the planning horizon. The supplies and demands of empty containers associated with ports reflect the fact that, in general, any transportation system is embedded into a larger one: regions into states, states into the world, etc. In particular, these values reflect the differences between supplies and demands at the level of the international trade. Let us then define:

- X_i^t : Demand of the identified customer i , who requires that containers be delivered at its site in period t , at the latest;
 $i \in I^t, t = 1, 2, \dots, T.$
- X_j^t : Export demand at port j in period t ; it represents the empty container requests which originate outside the study area;
 $j \in H, t = 1, 2, \dots, T.$
- Y_s^t : Supply of containers available in period t , at supply customer s ;
 $s \in S^t, t = 1, 2, \dots, T.$
- Y_j^t : Import supply at port j in period t ; it represents the input of empty containers from outside the study area;
 $j \in H, t = 1, 2, \dots, T.$

The Decision Variables

Generally speaking, the decision variables represent container movements through time and space. Those movements which involve customers are identified by the letter v , while w is used for all others.

In all formality, the variable which corresponds to the allocation decision from depot to customer should be noted

- $v_{ji}^{it'}$: Volume of containers allocated from depot j to customer i in period t , and arriving at destination in period t' ;
 $i \in I_j^t, j \in D \cup H, t = 1, 2, \dots, T.$

to clearly specify the spatial-temporal occurrence that represents the identified customer.

However, once the starting time period of a movement is specified, the arriving time is completely determined by the corresponding transit time, and hence it is **not a decision**. The same comments apply to all the decision variables defined therein, and we can thus use a much simpler notation.

We use the notation $\tau_{\eta\mu}$ to identify the *transit times* of the various movements, where the origin η and the destination μ may represent a depot, a customer or a port. These delays include the time required for loading at origin and unloading at destination, plus the necessary time to journey between these two points. We assume all transit times to be independent of their starting (t) and arriving (t') periods.

We need the following additional notation:

- I_j^t : Set of all demand customer requests which may be satisfied from depot j by a movement which starts in period t and arrives in period $t + \tau_{ji}$; i.e., $i \in I_j^t$ if $t' - \Delta_i \leq t + \tau_{ji} \leq t'$ for $i \in I^t$;
 $j \in D \cup H$, $t = 1, 2, \dots, T$.
- $J_i^{t'}$: Set of all depots which, in period t' , may be used to satisfy the demand customer request i , $i \in I^t$; the shipment starts in period t' and arrives in period $t' + \tau_{ji}$ such that $t - \Delta_i \leq t' + \tau_{ji} \leq t$; $J_i^{t'} \subseteq D \cup H$.
- J_s^t : Set of all depots where supply (customer) s may send, in period t , the empty containers present on-site; in particular, J_s^t includes $j(s)$, the port of origin of the loaded movement which generated the current empty supply;
 $J_s^t \subseteq D \cup H$, $s \in S^t$, $t = 1, 2, \dots, T$.
- $S_j^{t'}$: Set of all supply customers s , $s \in S^{t'}$, $t' = t - \tau_{sj}$, from which depot j may receive containers in period t ;
 $j \in D \cup H$, $t = 1, 2, \dots, T$.

We can now formally define the following decision variables:

- v_{ji}^t : Volume of containers allocated in period t , from depot j to customer i . The shipment arrives in period $t' = t + \tau_{ji}$;
 $j \in D \cup H$, $i \in I_j^t$, $t = 1, 2, \dots, T$.

- v_{sj}^t : Volume of empty containers picked up in period t at supply customer s , and arriving at depot j in period $t + \tau_{sj}$;
 $s \in S^t$, $j \in J_s^t$, $t = 1, 2, \dots, T$.
- w_j^t : Stock of empty containers available at depot j at the end of period t (and at the start of period $t + 1$);
 $j \in D \cup H$, $t = 1, 2, \dots, T$.
- w_{jk}^t : Flow of empty containers shipped from depot j to depot k . Shipment starts in period t and is completed in period $t + \tau_{jk}$;
 $j \in D \cup H$, $k \in D \cup H$, $t = 1, 2, \dots, T$.

The Constraints

We are now ready to write down the constraints of the deterministic, single commodity model.

Identified Customer Demand:

$$\sum_{t' \leq t} \sum_{j \in J_i^{t'} | \tau_{ji} = t - t'} v_{ji}^{t'} = X_i^t - K_i^t \text{ for all } i \in I^t, t = 1, 2, \dots, T. \quad (1)$$

Here, the demand of an identified customer may be satisfied from all the depots to which the customer is linked, by shipments which arrive in any of the periods that fall in the time interval defined by the customer's delivery window. In some cases, part of the identified demand may already have been satisfied by shipments initiated in periods prior to the beginning of the planning horizon (either because the customer's delivery window starts before this time, or because these shipments have yet to arrive at the customer's site). As these shipments correspond to known decisions, a constant value K_i^t is used in equation (1) to adjust the identified demand for the planning horizon (of course, the definition of the set $J_i^{t'}$ is modified accordingly):

- K_i^t : Volume of containers sent to customer i , $i \in I^t$, before the beginning of the current planning horizon.

Note that, to insure feasibility, we assume that for each identified customer demand i , $i \in I^t$, not all the sets $J_i^{t'}$ are empty. In operational terms, this implies that there is always a

depot sufficiently close by to satisfy the demand of an identified customer. This corresponds to the fact that the sales department of the company should not accept an order which is temporarily infeasible.

Customer Supply:

$$\sum_{j \in J_s^t} v_{sj}^t = Y_s^t \text{ for all } s \in S^t, t = 1, 2, \dots, T. \quad (2)$$

Empty containers, which are available at a customer site, have to be moved out, and may be sent either to the port where the initial loaded movement originated, or to one (or several) of the depots to which the customer is assigned in the strategic/tactical plan.

Stocks at Non-Port Depots:

$$\begin{aligned} w_j^t = & w_j^{(t-1)} + \sum_{t' \leq t} \left[\sum_{k \in D \cup H \mid \tau_{kj} = t - t'} w_{kj}^{t'} + \sum_{s \in S_j^t \mid \tau_{sj} = t - t'} v_{sj}^{t'} \right] + b_j^t + K_j^t \\ & - \sum_{k \in D \cup H} w_{jk}^t - \sum_{i \in I_j^t} v_{ji}^t \text{ for all } j \in D, t = 1, 2, \dots, T. \end{aligned} \quad (3)$$

For each depot, the stock of empty containers is computed at the end of each time period. Output of the depot are the volumes shipped to the other depots, to ports and to customers. The volume input during the period is made up of container traffic that left in previous periods from the other depots, the ports and the supply customers, and that is arriving to the depot in the current period, according to the corresponding transportation times. This volume is added up to the stock available at the beginning of the period, and to the containers which are brought in from outside the system:

b_j^t : Volume of containers "borrowed" (from partner companies; it may also represent rented containers, or newly bought ones) at depot j in period t ;
 $j \in D \cup H, t = 1, 2, \dots, T.$

We denote by w_j^0 the stock available at depot $j, j \in D \cup H$, at the beginning of the planning horizon. As previously, we use a constant to adjust the depot supply for the containers shipped to the depot before the beginning of the current planning horizon:

K_j^t : Volume of containers sent to depot j before the beginning of the current planning horizon and arriving in period t ;
 $j \in D \cup H$, $t = 1, 2, \dots, T$.

Port Supply, Demand and Stock:

$$w_j^t = w_j^{(t-1)} + \sum_{t' \leq t} \left[\sum_{k \in D \cup H \mid \tau_{kj} = t - t'} w_{kj}^{t'} + \sum_{s \in S_j' \mid \tau_{sj} = t - t'} v_{sj}^{t'} \right] + b_j^t + K_j^t \quad (4)$$

$$- \sum_{k \in D \cup H} w_{jk}^t - \sum_{i \in I_j} v_{ji}^t - X_j^t + Y_j^t + e_j^t \text{ for all } j \in H, t = 1, 2, \dots, T.$$

where

e_j^t : Unsatisfied demand of empty containers in port j , in period t ($e_j^t \geq 0$);
 $j \in H$, $t = 1, 2, \dots, T$.

The constraints determining the stock levels at port depots are very similar to those controlling stocks at non-port depots, with the exception that the net import/export demand at the port has to be met. To insure feasibility, since bounds may exist on the volumes that can be brought in, we introduce an *unsatisfied demand* variable, mainly to account for the impossibility in some cases of handling the export demand. This is the only demand we assume possible to reject. Note that the presence of both the b_j^t and e_j^t variables, with the corresponding costs, allows one to perform a number of interesting analyzes on the trade-offs between the costs of not meeting the export demand versus the cost of bringing in new containers into the system.

Massive Equilibration Flows:

$$\delta_{jk}^t l_{jk}^t \leq w_{jk}^t \leq \delta_{jk}^t u_{jk}^t \text{ for all } j \in D \cup H, k \in D \cup H, t = 1, 2, \dots, T. \quad (5)$$

$$\sum_{i \in \Delta} \delta_{jk}^t = 1 \text{ for all } j \in D \cup H, k \in D \cup H. \quad (6)$$

where

- δ_{jk}^t : equals 1 if a balancing movement may be initiated in period t , from depot j to depot k , and 0 otherwise;
 $j, k \in D \cup H, t = 1, 2, \dots, T.$
- $l_{jk}^t (u_{jk}^t)$: lower (upper) bound on a balancing movement of containers initiated in period t , from depot j to depot k ;
 $j, k \in D \cup H, t = 1, 2, \dots, T.$

The last two sets of equations insure that balancing movements are performed according to the rules and following the guidelines of the strategic/tactical plan: at specified periods and in consequent volumes. Several variants of equations (5) and (6) may be defined by suitably defining the set Δ . We further discuss this issue latter on.

The Objective Function

To write the objective function of the model, we need to define the "cost" criteria associated with each one of the transportation and warehousing activities defined in the system. We do not define any specific functional form yet, these are application dependent, but discuss cost related issues in Section 7.

In all generality, the cost function associated with a specific activity (e.g. moving containers, in a specific time period, between a given customer and a particular port) could depend upon the values of all flow variables in the model. Yet, in practice, such a broad definition is not required, since most empty traffic that involves customers takes place over relatively short distances and moves only a few containers at one time (this type of movement is almost always truck based). We then assume that for this type of activity, the unit cost function depends only on the intensity of the activity, and we define the following notation:

- $C_{ji}^t(v_{ji}^t)$: Transportation cost function from depot j to identified customer i , $i \in I'$, in period t ;
 $j \in D \cup H, t = 1, 2, \dots, T.$
- $C_{sj}^t(v_{sj}^t)$: Transportation cost function from supply customer s to depot j in period t ;
 $s \in S', j \in D \cup H, t = 1, 2, \dots, T.$

$C_{jk}^t(w_{jk}^t)$: Transportation cost function from depot j to depot k in period t ;
 $j \in D \cup H$, $k \in D \cup H$, $t = 1, 2, \dots, T$.

$C_j^t(w_j^t)$: Holding cost function at depot j in period t ;
 $j \in D \cup H$, $t = 1, 2, \dots, T$.

Finally, we introduce

$\bar{C}_j^t(b_j^t)$: The cost of bringing in containers from outside the system, at depot j , in period t ;
 $j \in D \cup H$, $t = 1, 2, \dots, T$.

$\tilde{C}_j^t(e_j^t)$: Penalty cost for not satisfying the external demand of empty containers at port j in period t ;
 $j \in H$, $t = 1, 2, \dots, T$.

The total cost of operating the land distribution and transportation system may then be written as:

$$\sum_{t=1,2,\dots,T} \left\{ \sum_{j \in D \cup H} \left[\sum_{i \in I_j^t} C_{ji}^t(v_{ji}^t) v_{ji}^t + \sum_{k \in D \cup H} C_{jk}^t(w_{jk}^t) w_{jk}^t + C_j^t(w_j^t) w_j^t + \bar{C}_j^t(b_j^t) b_j^t \right] \right. \quad (7)$$

$$\left. + \sum_{j \in H} \tilde{C}_j^t(e_j^t) e_j^t + \sum_{s \in S^t} \sum_{j \in J_s^t} C_{sj}^t(v_{sj}^t) v_{sj}^t \right\}$$

and the model minimizes (7), subject to constraints (1) to (6) and the usual non negativity restrictions on the decision variables.

3.2 The Network Model

The problem we study displays a natural network structure, which is present in the deterministic model of this section, as well as in the more complex formulations that follow. As a basis for these formulations, we introduce the minimum cost network flow model that corresponds to the single product deterministic case. The transformation involves the construction of the network and the specification, for each arc, of the flow variables, of the corresponding flow upper and lower bounds and of the cost function. The usual flow conservation constraints may be then directly inferred. However, when the network variables and constraints correspond rather straightforwardly to those of the previous model, we do not explicitly

rewrite them.

The formulation is based on a network $G = (N, A)$, whose spatial topology conceptually unfolds in two dimensional space: *time* spans the x axis, while *space (location)* unrolls on the y axis. The time axis is divided into periods, and for each period a network representation of the actual distribution and transportation system of the company unfolds on the space axis. This representation is specific to each particular period, since the sets of identified supply and demand customer requests usually are different for each time occurrence. Note that the parts of the network representations that correspond to the sets of ports and depots used by the company do not vary with the time period. "Dynamic" arcs link geographical representations along the time axis, and these arcs may span more than one time period, as will be shown below.

We illustrate the network model by using the simple case of Figure 2. The resulting network is displayed in Figure 3 for three time periods. It should be noted, however, that in order to keep the presentation clear and not overcharge the figure, we have drawn only "representative" arcs in Figure 3 instead of the complete network. Let us now detail this network representation.

We view the problem as a multisource multisink network. Hence, the node set N encompasses elements which represent the components of the distribution system: ports, depots and customers. The sets of nodes representing identified demand and supply customers are specific to each time period; hence, we use the I^t and S^t notation previously defined. The ports and depots used by the company being the same for each specific period, a replication of these elements is required for each time occurrence of the spatial representation. We therefore then define:

- H^t : Subset of nodes representing the replication of the set of ports in period $t, t = 1, 2, \dots, T$;
 $ht \in H^t$.
- D^t : Subset of nodes representing the replication in period $t, t = 1, 2, \dots, T$, of the non-port depots;
 $jt \in D^t, kt \in D^t$.

Finally, we add two super-source nodes to the node set N , to account for the previously defined relationships with the external world :

α : Source for the "new" (borrowed from partner companies, rented, bought, etc.) containers introduced into the system:

α' : Source for the external demand at ports that the system refuses to satisfy.

The definition and inclusion of the links of the network is mainly performed according to the company's operating policies, the strategic/tactical plan, the ship schedules and the temporal characteristics of supplies and demands. The plan specifies to which depot(s) each customer may be linked, and the depot pairs between which "massive" transfers of empty containers have to take place in order to insure a "globally optimal" equilibration of supply and demand in the system. The ship schedules directly determine when exchanges with the outside world occur and influence, through the forecasting process, when supplies and demands appear in the system. Additional arcs are included to model warehousing decisions, and to account for specific preoccupations of day-to-day operations.

Figure 3 displays part of the subnetwork which corresponds to three time periods of the distribution system described in Figure 2: one harbor and two inland depots, each replicated three times, plus two identified demand customer request nodes, it and $\hat{i}(t-1)$, and one identified supply customer request node $s(t-2)$. (When no risk of ambiguity exists, in the following we will only define the notation for period t .)

The arcs (ht, kt) and (kt, ht) are shown in order to illustrate the possibility of limited exchanges between "neighboring" depots. These exchanges require small transportation times and involve relatively limited volumes of containers. Such a movement may be initiated, for example, when the transportation cost from depot h to depot k plus the distribution cost starting from k is lower than the holding cost at h plus the corresponding distribution cost. It is important to note, however, that this type of movement is different from the massive inter-depot balancing traffic, represented by "dynamic" arcs of the type (jt, kt') (illustrated as arc $(j(t-2), ht)$ in Figure 3), although the flow variables are similarly identified. Lower, l'_{jk} , and upper, u'_{jk} , bounds usually control interdepot flows. We further discuss the utilization of the equilibration arcs, variables and link bounds in Section 3.3.

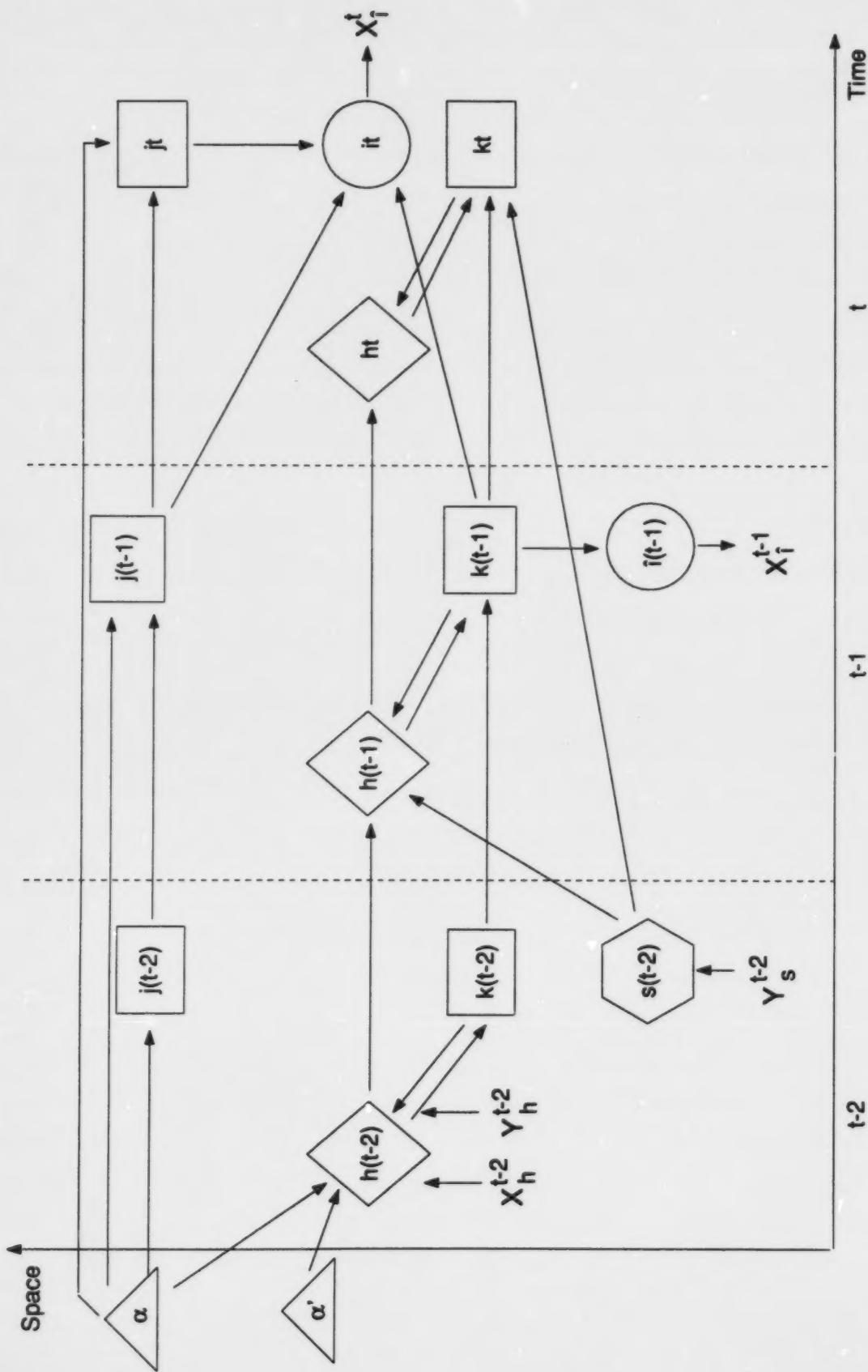


Figure 3. Outline of the Time-Space Network Representation

The drawn depot to customer arcs, of type (jt', it) (arcs $(k(t-1), \hat{i}(t-1))$, $(k(t-2), it)$, $(k(t-1), it)$ and (jt, it) of Figure 3), illustrate the fact that a customer demand may be met from several, close or not so close, depots. In particular, Figure 3 illustrates the "dynamic" arcs which are required to account for the delivery windows of identified demand customer requests (e.g. $(k(t-2), it)$ and $(k(t-1), it)$). Here, the arc lengths, measured as the number of periods spanned by the arcs, of the depot to customer links bear no relation to the transportation time τ_{ki} between the depot and the customer location. Rather, since it is assumed that one may always deliver during a customer's delivery window, they represent the decision to send a certain number of containers in period, say, $t-2$, in order to globally optimize the transportation and holding costs, while satisfying demand. The definition of these arcs has, however, to insure that one delivers in time, i.e. that $t-2 + \tau_{ki} \leq t$, for example. Thus, in Figure 3, the link (kt, it) has not been included since it would have violated the transportation time requirement.

We have also illustrated in Figure 3 some of the "holding" arcs, of type $(jt, j(t+1))$, which model the management of inventories of empty containers at depots. The upper limits, u_j^t , are usually derived from contractual limitations, or, more rarely, from actual restrictions. In both cases, these bounds are rather loose. The lower bounds, l_j^t , on the other hand, normally represent the safety stocks that the company evaluates, from past experience, that it should keep at each depot. The safety stock may vary with the depot location and, in some measure, with the time period.

Figure 3 also shows the two super-sources, α and α' . This enables us to illustrate the arcs used to model the various relationships with the outside world. Arc (α', ht) supports the flow, represented by variable e_j^t , required to satisfy the external demand for empty containers that the system cannot directly satisfy. Such arcs are introduced into the network only in periods when external demands are specified at port nodes. Arcs (α, ht) and (α, jt) capture the recourse to external sources of empty containers required at depots in order to satisfy the total demand. Although, in theory, the upper flow limit on these arcs should be equal to infinity (in order to insure feasibility) in practice some sensible and practical finite value bounds exist. Suitable total supplies have to be defined at nodes α and α' .

The arcs $(s(t-2), h(t-1))$ and $(s(t-2), kt)$ implement the requirement that empty containers have to be picked up at a (identified supply) customer's location in order to be transported either to the port of origin of the initial loaded movement, or to one (several) of the designated depots. Note that arcs between customers and depots span as many periods as required by their respective transportation times. Also note, that one may quite straightforwardly model the possibility that containers may be left at customers' sites for latter removal: one simply replicates the corresponding st nodes the number of periods the containers may be left there, the *customer holding* arcs $(st, s(t+1))$ capturing the effect of keeping inventories at customers' sites. Of course, appropriate cost functions have to be defined since, otherwise, significant volumes of containers might be left at customers' locations in order to avoid inventory costs at depots. This approach assumes that one does not need to pick up all containers present at a given location at the same time; to enforce this condition, one may formulate bundle constraints governed by binary variables, which would, however, significantly complicate the solution process. Hence, we rather propose, in this case, to use only one occurrence of the customer supply node, placed at the first period of availability if the cost of leaving containers at the customer site is higher than the average of the holding costs at the designated depots, and at the last possible period otherwise.

Finally, Figure 3 illustrates the demands and supplies associated with the nodes of the network. It may be noted that port supplies and demands appear only at those periods when ships are scheduled. Otherwise, the supplies and demands of the network model are the same, and appear at the same periods, that those of the previous formulation. In particular, supplies and demands are adjusted to account for the activities of previous periods, when necessary.

3.3 Notes and Comments

The following discussion is intended as a means of clarifying the formulation by an examination of the role played by the massive inter-depot balancing flows. It also hints at some of the additional complexity introduced when several container types are introduced.

Decisions concerning the initiation of an operation by which a significant volume of empty containers will be moved between two, normally rather distant, depots in order to equilibrate the regional (depot) supply and demand for some container types, are of a different nature than the other decisions considered in our modeling framework. Indeed, even if the

formulation aims at a global optimization of the distribution and transportation system, the issues addressed in this model concern, in general, the movement of a few containers (usually of the same type) at a time, over relatively short distances, by almost individual (trucks, generally) means of transportation. Hence, one is able to associate with most decisions an arc and a variable of the network model, which may be treated separately from the other arcs and variables. Moreover, these decisions are repetitive: the same operations have to be performed, although between different players, numerous times for each period, container type, even depot. Each decision, taken separately and myopically, has a limited influence on the state of the system; it is the combined effect of all these decisions that determines its behaviour.

Equilibration decisions are different. First, they concern larger volumes of containers of various types which travel together, over longer distances. Transportation is normally performed by modes efficient for this task, usually rail. Second, this type of decision occurs less frequently but has a sizable impact on the state of the system and may alter its behaviour. Hence, special care has to be taken in representing these decisions in the network formulation.

In the approach that we propose, all possible equilibration arcs exist, starting from all depots, in all periods. Binary variables (see equation (5) of Section 3.1) are then used to indicate which ones amongst them may be used at any given (*time, location*) occurrence. (In an actual implementation, only the permitted interdepot arcs would be included in the network.) The volume of containers implicated in the equilibration operation is then governed by upper and lower arc bounds handed down by the overall planning system from the strategic/tactical level.

Equations, such as equation (6) of Section 3.1, have then to be defined in order to enforce desired strategical/tactical decisions or traditional operational policies. For example, when $\Delta = \{1, 2, \dots, T\}$, equation (6) implies that equilibration traffic from depot j to depot k should occur only once over the planning horizon. Different definitions of the set Δ may be devised. Yet, none of these equations can insure that, in the multicommodity case for example, the following situation cannot happen: an equilibration flow for product p , from depot j to depot k , is ordered in period t , and another, for product r , is ordered in period $t+1$, while it takes the combination of the two volumes in order to achieve the critical mass required by massive interdepot equilibration flows. Hence, if the equilibration decisions are to be left to the allocation model, additional binary variables should govern the δ_{jk}^t binary variables which

rule the equilibration arcs. This would mean an almost unsolvable problem.

We rather take the view that equilibration is a higher planning level decision, mostly driven by the level of inventories at depots, which should direct the allocation process, and not the converse. Hence, we propose that, when the planning horizon is advanced, the inventory levels at depots be examined and, by using a transportation algorithm for example, the "optimal" equilibration movements be identified. Then, the corresponding arcs can be included into the network, in order to allow the correct computation of supplies. In this way, the δ'_{jk} , l'_{jk} and u'_{jk} are variables in the transportation procedure and parameters in the allocation model. It is noteworthy that various scenario analyses may be performed by varying the configuration of the equilibration arcs in the dynamic network, or by perturbing the values of the δ'_{jk} , l'_{jk} and u'_{jk} parameters. These specific preoccupations largely exceed, however, the scope of the present paper.

4. STOCHASTIC SINGLE COMMODITY FORMULATIONS

This section is dedicated to the development of a formulation of the dynamic single commodity case that takes into account the stochastic nature of some of its major components. We aim at a formulation that is both analytically tractable and numerically solvable, and that serves as a basis for the multicommodity model of Section 5.

Prior to the presentation of the formulations, we briefly discuss the characteristics of the stochastic elements of the problem. Recall that we consider demands and supplies to be stochastic, while the durations of the various transportation activities are assumed to be deterministically known. Also, in the dynamic context that we are considering, the planning horizon is divided into T periods, the first being "deterministic", its associated variables corresponding to decisions which may (have to) be implemented. The uncertainties thus affect only periods 2, 3, ..., T , and serve to ponder today's decisions by their estimated effect on future ones. How these periods are divided into stages depends on the particular formulation considered and is described alongside it. Finally, issues related to the end of the planning horizon are discussed later on, in Section 7.

4.1 The Stochastic Elements

The stochastic elements we consider are associated with the "internal", customer related, demands and supplies, as well as to the demands and supplies of empty containers from outside the system which are associated to exit and entry points, ports in particular.

Considering the internal demand, we distinguish two categories. The first one may be associated to a given customer, it may be described in physical and temporal terms (by using firm orders or historical data), and may be reasonably well forecast. As the exact volume requested is not generally known in advance, the demand is a random variable. The second category of customer related demand that we consider models the materialization of requests which are possible, and which should therefore be expected, but which are not, at this time, known with any certainty. We call this type of demand *diffuse*, and associate it to the depots.

The diffuse demand is made up of two basic items: (1) the demand of "regular" customers (i.e., customers who have already ordered containers of the current type many times during the year) which cannot yet be ascertained since no specific request has been placed and

since there is no clear cyclic pattern of orders, and (2) the demand of "spot" customers who do not usually deal with the company. As one goes further away on the temporal axis, the ratio of the diffuse demand to the total demand will thus tend to increase, since less and less is known about the customers' requests. The diffuse demand distribution is also used to capture the effect on the stock of containers of various other "catastrophic" events: accidents of all sorts implying that containers are scrapped or temporarily taken out of the system for repairs, requests for empty containers from partner shipping companies, etc.

Supplies of empty containers are also divided into two categories: those corresponding to *regular supply customers*, and the *diffuse* supply of containers available at depots. The diffuse depot supply stands for events which add up to the stock of empty containers, but which display a highly uncertain nature, such as the returns of repaired or lent containers, etc. On the other hand, for regular supply customer requests the uncertainty mainly derives not from the randomness of the amount of containers available for pick up, but rather from the "when does that supply become available" type of question.

Indeed, one has to note that supply customer requests are generated by the overall planning system from precise orders for, or even from actual, movements of a given number of loaded containers from an import port to the customer's location. Hence, since the temporary or permanent disablement of containers is modeled elsewhere, the volume of supply available at a customer's location is quite well known. In fact, but for the concern about the generality of the formulation, the amount of a customer supply should not be considered a random variable. Furthermore, if the request for shipping away empty containers corresponds to an actual customer call, then the temporal location of the request in the model is precisely known, and no stochasticity is really present.

In the general case, however, the availability of empty containers at customer sites is only forecast, in the sense that the planning system generates the request based on the knowledge that loaded containers are present, or are arriving, at a customer's location. Thus, in most cases, the moment when the empty containers will be available *becomes a random event*. Indeed, if loaded containers are delivered today, this does not necessarily mean that the containers will be emptied by, say, the following day. It may be tomorrow, or the day after, or the day after that. In fact one may infer, from historical data, the amount of variation particular to a given customer, variation that may increase when the loaded containers have

not yet arrived at their destination. Hence the customer supply becomes again a random variable, since the same request may appear at several, mutually exclusive, successive moments during the planning horizon.

Finally, the external supplies and demands of empty containers associated with ports may be also viewed as random variables which reflect, in particular, the differences between supplies and demands at the level of the international trade.

4.2 Assumptions and Possible Formulations

The deterministic model of Section 3 describes fairly extensively the complex relationships existing between the various decisions that have to be made when addressing the problem of optimally allocating containers over a large network. Ideally, we would like to be able to come up with a stochastic model which would account for the uncertain nature of supplies and demands, while retaining as much as possible of the interactions between the decision variables. At the same time, we would also like this model to be computationally tractable for large size networks. As we shall see further on, these objectives are unfortunately contradictory. The purpose of this section is to examine a series of possible formulations in which different trade-offs are made between the complexity of the interactions and the tractability of the models.

We first consider a direct adaptation of the model of the previous section. Without going into all the intricacies of defining exactly when the random variables describing demands and supplies are realized, let us simply suppose that new information on these becomes available at the beginning of each period. Under such an assumption, it should be possible to revise the allocation of containers at the beginning of each period, and the natural way to state the problem is as a *T-stage stochastic program with full network recourse* (this indicates that all future decisions are re-optimized in each period). Given the complex relationships linking the decision variables in different periods, the simplest way to write down this model is by resorting to a *scenario formulation*, in which each scenario corresponds to a specific joint realization of the random variables. Such a formulation yields a large mathematical program in which all decision variables and constraints have to be replicated for each scenario; these are linked by a set of *non-anticipativity constraints* which state that all decision variables in periods $1, 2, \dots, t$ must be equal for all scenarios in which random variables realized up to stage

t take on the same values. One may notice that, were it not for the non-anticipativity constraints, the overall problem would decompose in a series of smaller subproblems, each corresponding to a scenario (these subproblems are essentially deterministic in nature, since the random variables now take on specified values). Breton and El Hachem (1991) have taken advantage of this situation to propose an interesting solution scheme for this type of problem. This scheme is based on the application of Lagrangian relaxation to the non-anticipativity constraints. It involves performing a number of iterations consisting of three basic steps: in the first step, all subproblems are solved for fixed values of the Lagrange multipliers of the non-anticipativity constraints; in the second step, the decision variables corresponding to identical information subsets are combined by using a weighing procedure; in the third step, the multipliers are updated through subgradient optimization for the next iteration. Under suitable convexity assumptions, this procedure can be shown to converge to the optimal solution of the original problem. It is interesting to note that the subproblems solved at each iteration only differ by the addition to their objective function of linear "correction terms" which account for the changes in the Lagrange multipliers.

In the context of our problem, each scenario subproblem to be solved under the Breton-El Hachem approach corresponds to the deterministic model of Section 3 with a slightly modified objective function. This implies that if the number of possible joint realizations of the random demands and supplies were small, we could use this approach to solve the stochastic version of our problem. In general, however, this is not the case, as real-life instances typically display an extremely large number of possible realizations.

One way to somewhat simplify the previous model would be to assume that all the information regarding uncertain supplies and demands becomes available at the beginning of the second period (all random variables are thus realized at that time). This yields a *two-stage stochastic program with full network recourse* in which periods 2 to T are lumped together in the second stage. Such a simplification however does not significantly reduce the computational burden since each evaluation of the recourse function associated with the second stage, for a given set of values of the first period decision variables, still involves solving a deterministic problem with $(T - 1)$ periods for each possible realization of the decision variables.

If we want to achieve a tractable model, we must therefore accept much stricter restrictions regarding the recourse strategy. The approach that we have decided to use in this paper is that of *simple recourse*. Roughly stated, simple recourse implies partitioning the decision variables into two sets: *deterministic* decision variables (the "real" decision variables in a sense) which are determined by an optimization procedure on the basis of the information available at the beginning of the horizon (this information encompasses the parameters known with certainty as well as the forecasts on uncertain data), and *recourse* (random) variables which are defined to "adjust" the decisions made at the beginning of the horizon to the random events occurring in later periods as these unfold. All first period decision variables are deterministic.

The key element in a simple recourse approach is that recourse variables have to be evaluated without having to perform an explicit optimization procedure; the recourse variables are thus for all purposes direct functions of the random parameters. Simple recourse obviously limits the range of actions in future periods, but it allows us to account for uncertainties with a reasonable computational effort. It should also be pointed out that, if the model is properly defined, this approach should provide us with good conservative estimates of the costs to be incurred in the latter periods (we could certainly do better on the average by re-optimizing decisions on the basis of the new information which becomes available as time goes on). For these reasons, simple recourse models are best used in a *rolling-horizon* mode where only first-period decisions are implemented and the model is run anew at the beginning of each period.

Going back to the specific problem at hand, we must now carefully spell out how the random parameters will be defined and how the container allocation "system" will handle random events.

We start by clearly discriminating between deterministic and random events. In particular, we restrict the sets of identified customer requests I^t and S^t to deterministic occurrences only: the sets will contain only those requests for which everything is known with perfect certainty (i.e., quantities and delivery windows or periods of availability are already determined at the beginning of the planning horizon). For these events, the deterministic demands, supplies and flow variables defined in Section 3 will continue to be used.

With regards to uncertain requests, the case of the diffuse supplies and demands is easily handled since we have already assumed that these are associated with specific depots, for each time period, and that we have probability distributions to describe them. Uncertain "regular" customers, however, pose a difficult problem due to the fact that their requests may, in some cases, be satisfied from more than one depot. When such a request is realized, we should thus have to decide by which depots it will be handled. In the case of demands, this is further compounded by the necessity to make a decision on the timing of the container movements to the customer location (because of the flexibility afforded by the delivery windows). Such a latitude of action is hardly compatible with the requirements of a simple recourse model. To untangle the situation, we first force the assignment of each customer to a single depot for its uncertain requests of a given type (i.e., supply or demand) over the planning horizon. This is not, in general, such a big distortion, since most customers will already be assigned to a single depot for a given type of containers and requests as a result of planning at the strategic/tactical level. As for delivery windows, we can adopt, in a stochastic setting, the conservative policy of shipping containers to customers as late as possible, i.e., at the end of the forecast delivery windows. Having already forced the assignment of customers to a specified depot for uncertain demands, the period in which a shipment should occur under this policy can easily be determined.

Alternate hypotheses may be adopted, corresponding to different policies, as long as an uncertain supply or demand is associated to only one depot and one time period. However, once such an hypothesis is in effect, the demands and supplies of uncertain regular customers may be lumped together with the diffuse supplies and demands associated to depots. Hence, we define the random variables

- χ_j^t : Diffuse demand in the region of depot j , in period t ;
 $j \in D \cup H, t = 2, \dots, T.$
- ψ_j^t : Diffuse supply of containers available in period t at depot j ;
 $j \in D \cup H, t = 2, \dots, T.$

Deriving the probability distributions of these variables is a non-trivial task which may require extensive computations, since it will generally involve the convolution of several distributions defined for individual customers. However, this work is pre-processing with respect to the

resolution of the model, its description clearly being outside the scope of this paper.

Apart from customer requests, containers are received at ports from outside the system and export demands for empty containers are also made at these locations. To simplify the exposition, we assume that, except for the first period in which such supplies and demands should be known with certainty (we retain the notations X_j^1 and Y_j^1 , for $j \in H$, for these), these are random variables, denoted

ξ_j^t : Export demand at port j in period t ;
 $j \in H, t = 2, \dots, T.$

ϕ_j^t : Import supply at port j in period t ;
 $j \in H, t = 2, \dots, T.$

Having spelled out the random parameters describing uncertain events in our empty container distribution system, we may now define the recourse strategy. It must first be noticed that uncertain supplies are easily handled, since they just "appear" at depots and ports where they can be added to the stock on hand. With regards to uncertain customer demands, we have already stated that all customer demands must be satisfied. For a given depot j , $j \in D \cup H$, and time period t , $t = 2, \dots, T$, this implies that when χ_j^t is realized, we either satisfy it from available stocks (after taking into account all previously decided movements to other customers and depots), or we resort to bringing in "new" containers (which should be avoided whenever possible). Export demands at ports are different, since we have allowed them not be satisfied in the deterministic model. For simplicity sake, we will assume in the stochastic model that, except for period 1 where this is a deterministic decision, they will be satisfied up to the level of available stocks.

The overall operations of the container distribution system encompass not only the movements of containers to and from customers, but also those between depots. We could eventually make depot-to-depot movements in periods 2 to T dependent upon the actual stocks of empty containers at depots, but this would immensely complicate the formulation of the model. For this reason, we prefer at this stage of our research to make these decision variables deterministic (all such flows are thus determined at the beginning of the planning horizon), even though, in some cases, this may lead to strange situations such as having to borrow containers from partner companies in order to be able to implement balancing movements; it can

be expected, however, that the optimization procedure will minimize such occurrences in the optimal solution of the model.

To summarize the previous discussion, there will be three types of recourse variables in the model: the stocks at port and non-port depots, the quantities of "new" containers (bought, borrowed from partner companies, rented from elsewhere, etc.) and the unsatisfied export demand at ports.

Before closing this section and proceeding to the formal description of the stochastic model, two remarks are in order. First, it should be noticed that we have lost the exact location of customer sites by performing the aggregation of uncertain demands and supplies. It is thus no longer possible to compute the exact cost of movements between these sites and their assigned depots. But since these movements are "forced" in the model, they have no impact on the optimization; they can therefore be dropped from the objective function. Second, readers familiar with stochastic programming will have noticed that the recourse structure that we have defined is not the "classical" simple recourse structure with overage and underage variables. In fact, it is quite similar to the *null recourse* strategy introduced by Powell(1988).

4.3 The Stochastic Model

In the description of this model, as well as of the general stochastic model of the next section, we distinguish between deterministic quantities (parameters and decision variables), and random variables (parameters and recourse variables), by using Latin letters for the first group and Greek letters for the second.

The decision variables of the one commodity stochastic model are, for the most part, the same as defined in the previous section for the deterministic formulation. We need, however, to define the *recourse* variables: the stocks at ports and depots, the recourse to external sources of empty containers at depots, and the unsatisfied export demand at ports. Let

ω_j^t : Stock of empty containers available at depot j at the end of period t ;
 $j \in D \cup H, t = 2, \dots, T.$

β_j^t : Volume of "new" containers at depot j in period t ;
 $j \in D \cup H, t = 2, \dots, T.$

ϵ_j^t : Unsatisfied demand of empty containers in port j , in period t ;
 $j \in H, t = 2, \dots, T$.

Using this notation, and the constants K_i^t and K_j^t previously defined to represent containers sent to customers, depots and ports before the beginning of the planning horizon and arriving in the current period, we may now write down the constraints of the model. These can be partitioned into two groups:

- i) Constraints linking the deterministic variables and thus defining their feasible sets; these constraints are similar to those of the deterministic formulation: equations (1), (2), (5) and (6) of Section 3, plus

Stocks at Non-Port Depots in Period 1 :

$$w_j^1 = w_j^0 + b_j^1 + \sum_{k \in D \cup H \mid \tau_{kj} = 0} w_{kj}^1 + K_j^1 - \sum_{k \in D \cup H} w_{jk}^1 - \sum_{i \in I_j^1} v_{ji}^1 \quad (8)$$

for all $j \in D$.

Port Supply, Demand and Stock in Period 1 :

$$w_j^1 = w_j^0 + b_j^1 + \sum_{k \in D \cup H \mid \tau_{kj} = 0} w_{kj}^1 + K_j^1 - \sum_{k \in D \cup H} w_{jk}^1 - \sum_{i \in I_j^1} v_{ji}^1 - X_j^1 + Y_j^1 + e_j^1 \quad (9)$$

for all $j \in H$.

- ii) Definitional constraints for the recourse variables in periods 2, ..., T , which allow us to compute the expected recourse for any given values of the deterministic variables.
 Define

σ_j^t : The net balance of containers at depot j in period t before uncertain demands are taken into account;
 $j \in D \cup H, t = 2, \dots, T$.

We then have

Stocks at Non-Port Depots in Subsequent Periods:

$$\begin{aligned} \sigma_j^t = & \omega_j^{(t-1)} + \psi_j^t + \sum_{t' \leq t} \left[\sum_{k \in D \cup H \mid \tau_{kj} = t - t'} w_{kj}^{t'} + \sum_{s \in S_j' \mid \tau_{sj} = t - t'} v_{sj}^{t'} \right] \\ & + K_j^t - \sum_{k \in D \cup H} w_{jk}^t - \sum_{i \in I_j} v_{ji}^t \quad \text{for all } j \in D, \quad t = 2, \dots, T. \end{aligned} \quad (10)$$

Equation (10) states that the net balance of empty containers at depot j , $j \in D$, is the sum of the net balance of the deterministic container flows, the previous (random) inventory, plus the (random) uncertain supply at the depot. This balance may be either positive or negative. Note that, for $t = 2$, $\omega_j^{(t-1)} = w_j^1$.

If we denote by $[x]^+$ the positive part of x (i.e., $\max(0, x)$), the net inventory at depot j , $j \in D$, in period t , $t = 2, \dots, T$, is given by:

$$\omega_j^t = [\sigma_j^t - \chi_j^t]^+, \quad (11)$$

and the number of containers that will have to be brought in from outside sources to balance the flows is

$$\beta_j^t = [\chi_j^t - \sigma_j^t]^+. \quad (12)$$

Similarly, we have

Stocks at Ports in Subsequent Periods:

$$\begin{aligned} \sigma_j^t = & \omega_j^{(t-1)} + \psi_j^t + \phi_j^t + \sum_{t' \leq t} \left[\sum_{k \in D \cup H \mid \tau_{kj} = t - t'} w_{kj}^{t'} + \sum_{s \in S_j' \mid \tau_{sj} = t - t'} v_{sj}^{t'} \right] \\ & + K_j^t - \sum_{k \in D \cup H} w_{jk}^t - \sum_{i \in I_j} v_{ji}^t \quad \text{for all } j \in D, \quad t = 2, \dots, T. \end{aligned} \quad (13)$$

Since customer demands are always satisfied but export demands are met only if containers are on hand, we have the following relationships for the stocks, "new" containers, and unsatisfied export demands at port j , $j \in H$, in period t , $t = 2, \dots, T$:

$$\omega_j^t = [\sigma_j^t - \chi_j^t - \xi_j^t]^+, \quad (14)$$

$$\beta_j^t = [\chi_j^t - \sigma_j^t]^+, \quad (15)$$

and

$$\epsilon_j^t = \min \{ \xi_j^t, [\xi_j^t + \chi_j^t - \sigma_j^t]^+ \}. \quad (16)$$

A few words of explanation regarding these constraints. Equations (11) and (14) state that the inventory at the end of period t will be positive only if there are containers left over after satisfying both customer and, for ports, export demands, while equations (12) and (15) indicate that containers will only be borrowed (or bought) to cover a deficit created by deterministic movements (this would occur if $\sigma_j^t < 0$) and uncertain customer demands. Finally, it can easily be verified that equation (16) implies that

$$\epsilon_j^t = \xi_j^t \quad \text{if } \beta_j^t > 0$$

$$\epsilon_j^t = 0 \quad \text{if } \omega_j^t > 0$$

and

$$\epsilon_j^t = \xi_j^t + \chi_j^t - \sigma_j^t \quad \text{if } \beta_j^t = \omega_j^t = 0.$$

The objective function of the stochastic model is now to minimize the total expected cost of the empty container distribution system which, using the cost functions defined in Section 3, may be written as

$$\begin{aligned} & \sum_{t=1,2,\dots,T} \left[\sum_{j \in D \cup H} \left[\sum_{i \in I_j^t} C_{ji}^t(v_{ji}^t) v_{ji}^t + \sum_{k \in D \cup H} C_{jk}^t(w_{jk}^t) w_{jk}^t + \sum_{s \in S^t} \sum_{j \in J_s^t} C_{sj}^t(v_{sj}^t) v_{sj}^t \right] \right. \\ & + \sum_{j \in D \cup H} \left[C_j^1(w_j^1) w_j^1 + \bar{C}_j^1(b_j^1) b_j^1 \right] + \sum_{j \in H} \bar{C}_j^1(e_j^1) e_j^1 \\ & + E_{\chi, \psi, \xi, \phi} \left[\sum_{t=2,\dots,T} \left[\sum_{j \in D \cup H} \left[C_j^t(\omega_j^t) \omega_j^t + \bar{C}_j^t(\beta_j^t) \beta_j^t \right] + \sum_{j \in H} \bar{C}_j^t(\epsilon_j^t) \epsilon_j^t \right] \mid \underline{v}, \underline{w} \right] \end{aligned} \quad (17)$$

where $\underline{v}$ and $\underline{w}$ represent vectors of deterministic flow variables.

5. STOCHASTIC MULTICOMMODITY FORMULATIONS

The possibility to substitute one type of container for another increases the flexibility of the distribution system and enhances its ability to adequately respond to the customers' requests. Hence, our final and most general model is a multicommodity formulation, where each container type is represented by a different commodity or product, and where partial substitutions are allowed along the various container types. Note that, in a stochastic context, substitutions represent an additional mechanism to react to fluctuations in supplies and demands. The multicommodity formulation thus presents a richer recourse strategy than the single commodity one.

To make for a clearer presentation, we first describe the modeling of the multicommodity case in a deterministic setting. The stochastic formulation comes next.

5.1 The Deterministic Model

Let P be the set of container types (products), with $p \in P$. Then, in general, the notation of the preceding sections generalizes rather straightforwardly to the multicommodity case by inserting the product index p before the time index t . Hence, for example, the identified customer demand becomes:

$$\begin{aligned} X_i^{pt} : & \text{Demand of identified customer } i, \text{ who requires that containers of type } p \text{ be} \\ & \text{delivered at its site in period } t \text{ at the latest;} \\ & i \in I^{pt}, \quad p \in P, \quad t = 1, 2, \dots, T, \end{aligned}$$

while the corresponding node set is written I^{pt} .

For the sake of brevity, we do not explicitly redefine all variables, constants and sets, nor do we write down all the equations previously described. We indicate hereafter only those elements that change or that are new due to the multiproduct nature of the problem.

The network $G = (N, A)$ that underlies the deterministic formulation is now a three dimensional structure: *time* on the x axis, *space (location)* on the y axis, and *products (container types)* on the z axis. The product axis has as many units as the number of container types defined in the problem. The network representation of the actual distribution and transportation system of the company, which unfolds on the space axis, is now specific to each

particular (*time, product*) occurrence, since the sets of supply and identified demand customer requests usually vary with the container type as well as with the time period. Note that the sets of ports and depots used by the company continue to represent the invariant part of the network. "Dynamic" arcs link geographical representations along the time axis, while substitution arcs link the various commodity networks for a given time period.

To adequately represent the substitution mechanism and its consequences, we distinguish the containers which "physically" are of a given type p , from the *equivalent containers of type p* which are to be distributed to customers. This is tantamount to differentiating the two main depot functions: managing stocks of containers of various types and of allocating containers to satisfy customer demands. As a modeling principle, substitutions occur only at depots. Furthermore, stocks of containers at depots, all interdepot traffic, as well as all exchanges with the outside world, are of true containers of the given type, while demands are satisfied by using equivalent containers.

The customer allocation decision variable then becomes

$$v_{ji}^{pt} : \text{Volume of equivalent containers of type } p \text{ allocated in period } t \text{ from depot } j \text{ to customer } i;$$

$$j \in D \cup H, i \in I_j^p, p \in P, t = 1, 2, \dots, T;$$

while new decision variables have to be defined:

$$v_j^{pt} : \text{Total volume of containers of type } p, \text{ in stock at depot } j, \text{ which will be allocated, in period } t, \text{ as equivalent containers of type } p;$$

$$j \in D \cup H, p \in P, t = 1, 2, \dots, T;$$

$$w_j^{prt} : \text{Number of empty containers of type } p \text{ substituted for containers of type } r, \text{ at depot } j \text{ in period } t;$$

$$j \in D \cup H, p \in P, r \in P, t = 1, 2, \dots, T.$$

Substitution factors are used to represent some of the technical restrictions related to using a container of a type different from the one required by the customer. For any given pair (p, r) of products, the substitution factor a_{pr} represents the fact that more than one container of type p might be required to replace one container of type r . It should be stressed that these factors do not correspond to direct volume relationships between container types.

For instance, a 40-foot container is not equivalent to two 20-foot containers, since if a customer requires a single 20-foot container, a full 40-foot container would be needed to replace it (one can hardly think of a half-container moving around). On the other hand, two 20-foot containers could, in some circumstances, be substituted for a 40-foot one, but not always. The substitution factors (they could vary by depot to model particular conditions in some areas) are thus always the inverse of some integer number, such as

$$a_{pr} = (\text{number of type } p \text{ containers needed to replace one type } r \text{ container})^{-1}.$$

In the network formulation, this approach is implemented by the association of two nodes to each depot and port:

D^{pt} (H^{pt}) : Subset of nodes representing the replication, for container type p and period t , of the physical locations (depots and ports) where stocks of empty containers are managed;
 $p \in P, t = 1, 2, \dots, T.$

$\tilde{D}^{pt}$ ($\tilde{H}^{pt}$) : Subset of nodes representing the replication, for product p and period t , of the source depots (ports), from where equivalent empty containers of type p are allocated to customers;
 $p \in P, t = 1, 2, \dots, T.$

Then, for each container type layer p in the network, all flows between nodes in H^{pt} (the harbor sets), D^{pt} (depots) and S^{pt} (supply customer requests), for all periods t , represent movements of true containers of type p , as do flows between nodes jpt , in $D^{pt} \cup H^{pt}$ and their $j'pt$ replications in $\tilde{D}^{pt} \cup \tilde{H}^{pt}$. On the other hand, flows that originate from the second depot sets $\tilde{D}^{pt} \cup \tilde{H}^{pt}$ should be interpreted as movements of equivalent containers. This definition of the network structure ensures that containers that are substituted are directly used to satisfy demands and do not end up in inventories at depots. Figure 4 illustrates these concepts. For clarity reasons, it only displays part of the subnetwork that corresponds to the space-time layer of product p . The example network of Section 3 is also used here.

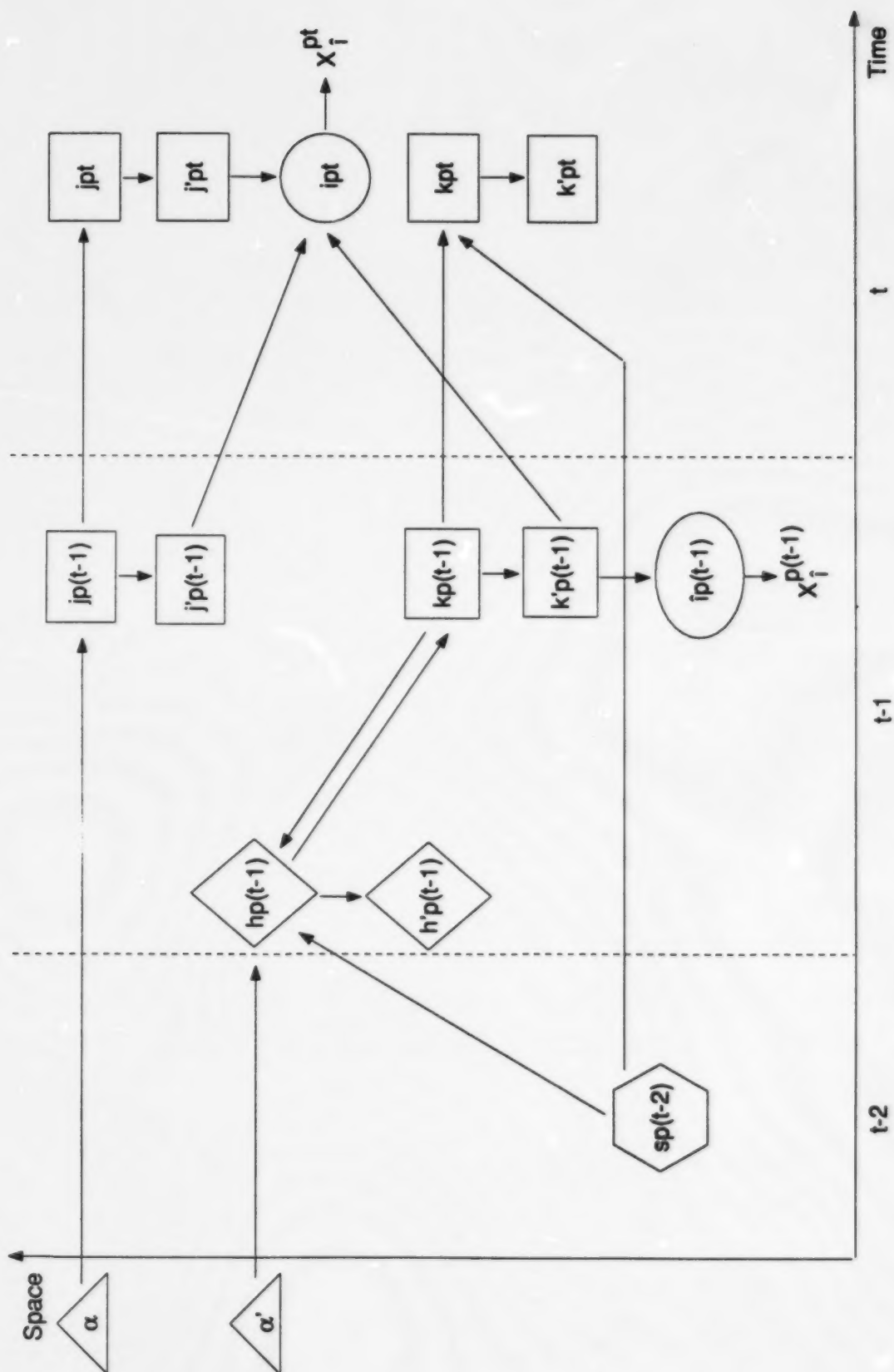


Figure 4. Outline of a Time-Space Container Type Layer

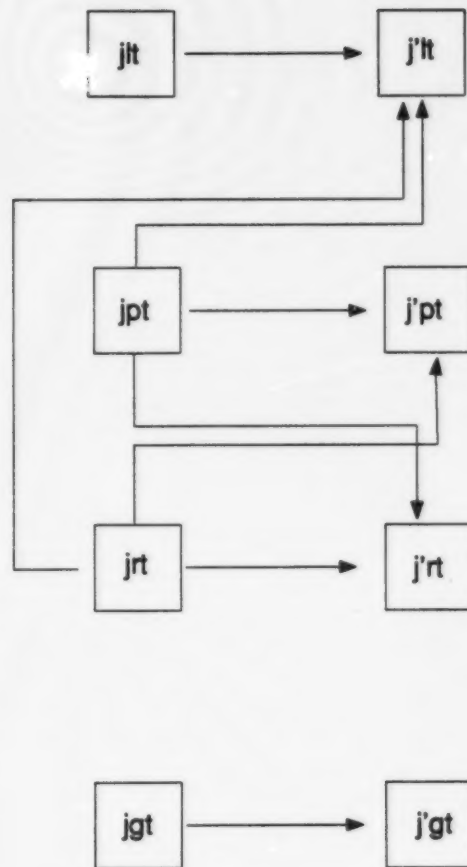


Figure 5. Partial Substitution Subnetwork

Movements originating from supply customer request nodes (S^{pt}) warrant a special note. Indeed, technically, these should be interpreted as representing traffic of "equivalent" containers. However, this option would rapidly put the system into a chaotic state, as "equivalent" and "physical" containers are mixed into the depot (nodes jpt) inventories. To avoid this, one should track each container individually, which would result into a cumbersome, unsolvable model. An elegant way of addressing this issue is based on the observation that supply customer requests are generated from actual movement orders for loaded containers. Consequently, when supplies are estimated, one knows exactly the container types involved. Hence, at generation time, the supplies at a customer's site may be transformed into as many supply customer requests as the number of different container types present. In this manner, the flows originating from nodes in S^{pt} represent "physical" container traffic.

Permitted substitution activities are represented by arcs linking different container type layers of the network. Figure 5 shows a number of $(jpt, j'rt)$ arcs which illustrate some of the possible substitution cases: on one hand, containers of type p and r may be substituted one for the other; on the other hand, containers of type p and r may be substituted for the type l , but the converse operation is not allowed. Finally, no substitutions are permitted between containers of type g and any of the types p, r or l . Flow variables w_j^{prt} measure the volume of substitutions, which are governed by the substitution (gain) factors a_{pr} . Upper bounds B_j^{prt} limit the amount of flow transferring from one commodity layer to another, thus modeling the actual restrictions present in company operating policies.

Relatively few modifications have to be performed to the constraints of the single commodity model, besides the insertion of the product index p . Indeed, only constraints (3) and (4) have to be modified to account for the containers that are used for substitutions

$$w_j^{pt} = w_j^{p(t-1)} + \sum_{t' \leq t} \left[\sum_{k \in D \cup H \mid \tau_{kj} = t - t'} w_{kj}^{pt'} + \sum_{s \in S_j^{pt'} \mid \tau_{sj} = t - t'} v_{sj}^{pt'} \right] \quad (18)$$

$$+ b_j^{pt} + K_j^{pt} - \sum_{k \in D \cup H} w_{jk}^{pt} - v_j^{pt} - \sum_{r \in P} w_j^{prt}$$

for all $j \in D, p \in P, t = 1, 2, \dots, T,$

and

$$\begin{aligned}
 w_j^{pt} = w_j^{p(t-1)} + \sum_{t' \leq t} \left[\sum_{k \in D \cup H \mid \tau_{kj} = t - t'} w_{kj}^{pt'} + \sum_{s \in S_j^{pt'} \mid \tau_{sj} = t - t'} v_{sj}^{pt'} \right] \\
 + b_j^{pt} + K_j^{pt} - \sum_{k \in D \cup H} w_{jk}^{pt} - v_j^{pt} - \sum_{r \in P} w_j^{prt} - X_j^{pt} + Y_j^{pt} + e_j^{pt} \\
 \text{for all } j \in H, p \in P, t = 1, 2, \dots, T,
 \end{aligned} \tag{19}$$

respectively, while a new set of constraints has to be defined to represent the

Depot Allocation Balance:

$$v_j^{pt} + \sum_{r \in P} a_{rp} w_j^{prt} - \sum_{i \in I_j^p} v_{ji}^{pt} = 0 \quad \text{for all } j \in D \cup H, p \in P, t = 1, 2, \dots, T. \tag{20}$$

Most of the cost definitions of the previous sections readily generalize to the multiproduct case. However, transportation activities between depots, as well as between depots and ports, may now involve containers of different types, due to possible consolidation. Similarly, warehousing costs may depend upon the total volume of containers in stock at a given depot. Hence, the corresponding cost functions will respectively depend on the

$\underline{w}_{jk}^t$: Vector of flows from depot j to depot k in period t , for all container types;

$$\underline{w}_{jk}^t = (w_{jk}^{pt}), \quad j \in D \cup H, k \in D \cup H, p \in P, t = 1, 2, \dots, T.$$

$\underline{w}_j^t$: Vector of the total stock of containers, of all types, kept at depot j at the end of period t ;

$$\underline{w}_j^t = (w_j^{pt}), \quad j \in D \cup H, p \in P, t = 1, 2, \dots, T.$$

On the other hand, substitution costs are assumed to depend only on the container types involved, and we define:

$C_j^{prt}(w_j^{prt})$: Substitution cost function, at depot j in period t , when containers of type p replace containers of type r ;
 $j \in D \cup H, p \in P, r \in P, t = 1, 2, \dots, T.$

To be perfectly correct, the depot to customer transportation cost function, $C_{sj}^{pt}(v_{sj}^{pt})$, should reflect the fact that, due to substitutions, the model allocates "equivalent" containers

to customers. This would imply, however, either that containers are individually tracked when allocated to customers, or that approximations based on the number of substitute containers are used. The first alternative is quite contrary to the spirit of the allocation model, and requires a different model structure, which, in effect, corresponds to the integration of the allocation and routing formulations. This is not, however, practical. The second alternative requires elaborate calibration procedures, thus losing in model generality. Furthermore, since the depot to customer cost functions would have to consider the incoming flows on the substitution arcs, this option brings substantial algorithmic difficulties due to the non arc-separability of some of the cost functions. We therefore prefer to address this phenomenon by introducing a corrective factor, based on historical data, in the substitution cost function. This approach dispenses with all the mentioned difficulties, while still allowing to adequately treat the problem.

The deterministic multiproduct model then minimizes

$$\begin{aligned} \sum_{i=1,2,\dots,T} \sum_{p \in P} \left[\sum_{j \in D \cup H} \left[\sum_{i \in I_j^p} C_{ji}^{pi}(v_{ji}^{pi}) v_{ji}^{pi} + \sum_{k \in D \cup H} C_{jk}^{pi}(w_{jk}^i) w_{jk}^{pi} \right. \right. \\ \left. \left. + C_j^{pi}(w_j^i) w_j^{pi} + \bar{C}_j^{pi}(b_j^{pi}) b_j^{pi} + \sum_{r \in P} C_j^{pri}(w_j^{pri}) w_j^{pri} \right] \right. \\ \left. + \sum_{j \in H} \bar{C}_j^{pi}(e_j^{pi}) e_j^{pi} + \sum_{s \in S^p} \sum_{j \in J_s^p} C_{sj}^{pi}(v_{sj}^{pi}) v_{sj}^{pi} \right] \end{aligned} \quad (21)$$

subject to constraints (1), (2), (18), (19), (20), (5) and (6), to upper bounds on the substitution variables, and the usual non negativity restrictions on the decision variables.

5.2 The Stochastic Formulation

As in the case of the deterministic model, most definitions of the single commodity stochastic model of Section 4.3 can readily be extended to the situation where multiple container types exist by simply adding a superscript p to variable and parameter names. The central question in the definition of the multicommodity stochastic model is, however, the handling of substitutions.

Obviously, substitutions in the first period, which is completely deterministic, can be treated exactly as in the deterministic model that we just described. It would be extremely tempting to limit substitutions to the first period only, in order to keep the model fairly simple. In our view, this would be a major error, since, in fact, substitutions are primarily used to deal with unforeseen stockouts of container types which are requested by customers. Substitutions are thus a widely used recourse action in the day-to-day operations of a container shipping company (in most instances, this mechanism would probably be preferred to borrowing from partner companies), and they should thus be allowed in all periods.

It should be obvious that the determination of the optimal substitution pattern at a given depot for any given period effectively involves a solving generalized transportation problem, which is clearly incompatible with the simple recourse structure that we have imposed in the single commodity model. In order to allow for substitutions in all periods while retaining the conceptual simplicity of the earlier model, we propose instead to use a *fixed-priority recourse* structure: a set of priority rules defining in which order possible substitutions should be handled. Such rules could be determined on the basis of company policies regarding substitutions. In the case of ports with possible export demands, these rules also specify what priorities should be enforced when situations arise where containers could be used to either satisfy export demands or replace other container types. Note that in all situations, we suppose that containers will first be allocated to their own type demands. We must also be careful not to use substitute containers to cover deficits created by container movements between depots since these must be "physical" container flows as in the deterministic model (substituted containers may only be allocated to customer demands).

To see how this fixed-priority recourse structure works, let χ_j^{pt} , ψ_j^{pt} , ξ_j^{pt} , ϕ_j^{pt} , ω_j^{pt} , β_j^{pt} and ε_j^{pt} , for all $t = 2, \dots, T$, be defined as in Section 4.3 with the superscript p added. Also, define

$\bar{w}_j^{pt}$: Net input (output, if negative) of containers of type p at depot j in period t arising from deterministic flows, excluding the identified customer demands;
 $p \in P$, $j \in D \cup H$, $t = 2, \dots, T$;

$\bar{v}_j^{pt}$: Total volume of equivalent containers of type p allocated to identified customers from depot j in period t ;

$$\bar{v}_j^{pt} = \sum_{i \in I^m} v_{ji}^{pt} \text{ for all } j \in D \cup H, p \in P, t = 2, \dots, T;$$

ω_j^{prt} : Number of containers of type p substituted for containers of type r , at depot j in period t ;

$$p \in P, r \in P, j \in D \cup H, t = 2, \dots, T;$$

L_j^t : The substitution priority structure for depot j at period t ;

$$j \in D \cup H, t = 2, \dots, T.$$

In most cases, one would expect to have identical priority lists in all periods for all ports, on one hand, and for non-port depots, on the other hand (i.e., $L_j^t = L_H$ for all $j \in H$ and $L_j^t = L_D$ for all $j \in D$, for all t). When $j \in D$, L_j^t is an ordered list of pairs (p, r) indicating possible substitutions of containers of type p for containers of type r (with substitution factor a_{pr} and upper bound B_j^{prt}); when $j \in H$, L_j^t also contains singletons $\{p\}$ for export demands of type p .

For each depot $j, j \in D \cup H$, and product $p, p \in P$, we define, for all periods $t, t = 2, \dots, T$, the following auxiliary random variables:

$\bar{\sigma}_j^{pt}$: Net balance of "physical" containers of type p at depot j in period t , before considering demands;

$$p \in P, j \in D \cup H, t = 2, \dots, T.$$

σ_{jq}^{pt} : Net balance of "equivalent" containers of type p at depot j in period t , excluding any deficit of "physical" containers, once the q first priority rules have been applied; σ_{j0}^{pt} is the net balance before applying any rule;

$$p \in P, j \in D \cup H, t = 2, \dots, T, q = 0, 1, \dots, |L_j^t|.$$

We first compute the balance of "physical" containers before considering demands. This is done in two steps. First, one considers only the deterministic flows

$$\bar{w}_j^{pt} = \sum_{t' \leq t} \left[\sum_{k \in D \cup H \mid \tau_{kj} = t - t'} w_{kj}^{pt'} + \sum_{s \in S_j^{pt'} \mid \tau_{sj} = t - t'} v_{sj}^{pt'} \right] - \sum_{k \in D \cup H} w_{jk}^{pt} + K_j^{pt} \quad (22)$$

$$\text{for all } j \in D \cup H, p \in P, t = 2, \dots, T,$$

and then by adding the random terms

$$\bar{\sigma}_j^{pt} = \bar{w}_j^{pt} + \omega_j^{p(t-1)} + \psi_j^{pt} \quad \text{for all } j \in D, p \in P, t = 2, \dots, T, \quad (23)$$

and

$$\bar{\sigma}_j^{pt} = \bar{w}_j^{pt} + \omega_j^{p(t-1)} + \psi_j^{pt} + \phi_j^{pt} \quad \text{for all } j \in H, p \in P, t = 2, \dots, T. \quad (24)$$

Note that, when $t = 2$, $\omega_j^{p(t-1)}$ is replaced by w_j^{p1} . Also, if $\bar{\sigma}_j^{pt} < 0$, we have a deficit in "physical" containers which will have to be satisfied by bringing in containers from outside the system (new ones, borrowed from partner companies, etc.).

We may now determine the net balance of "equivalent" containers, before applying any substitutions or considering export demands:

$$\sigma_{j0}^{pt} = [\bar{\sigma}_j^{pt}]^+ - \bar{v}_j^{pt} - \chi_j^{pt} \quad \text{for all } j \in D, p \in P, t = 2, \dots, T. \quad (25)$$

We can start applying the priority rules. If the q th rule is (p^*, r^*) , then for each depot $j, j \in D \cup H$, and time period $t, t = 2, \dots, T$, we have:

$$\sigma_{jq}^{pt} = \sigma_{j(q-1)}^{pt}, \quad \text{for all } p \neq p^*, r^*, \quad (26)$$

$$\omega_j^{p^*r^*t} = \left[\min \{ \sigma_{j(q-1)}^{p^*t}, (a_{p^*r^*})^{-1} (-\sigma_{j(q-1)}^{r^*t}), B_j^{p^*r^*t} \} \right], \quad (27)$$

$$\sigma_{jq}^{p^*t} = \sigma_{j(q-1)}^{p^*t} - \omega_j^{p^*r^*t}, \quad (28)$$

$$\sigma_{jq}^{r^*t} = \sigma_{j(q-1)}^{r^*t} + a_{p^*r^*} \omega_j^{p^*r^*t}. \quad (29)$$

If the q th rule is $\{p^*\}$ then, for each port $j, j \in H$, and time period $t, t = 2, \dots, T$, we compute:

$$\sigma_{jq}^{pt} = \sigma_{j(q-1)}^{pt}, \quad \text{for all } p \neq p^*, \quad (30)$$

$$\varepsilon_j^{p^*t} = [\min \{ \xi_j^{p^*t}, \xi_j^{p^*t} - \sigma_{j(q-1)}^{p^*t} \}]^+, \quad (31)$$

$$\sigma_{jq}^{p^*t} = \sigma_{j(q-1)}^{p^*t} - (\xi_j^{p^*t} - \varepsilon_j^{p^*t}). \quad (32)$$

Finally, we set

$$\omega_j^{pt} = [\sigma_{j\mu}^{pt}]^+ \quad \text{for all } j \in D \cup H, p \in P, t = 2, \dots, T, \quad (33)$$

$$\beta_j^{pt} = [-\bar{\sigma}_j^{pt}]^+ + [-\sigma_{j\mu}^{pt}]^+ \quad \text{for all } j \in D \cup H, p \in P, t = 2, \dots, T, \quad (34)$$

and

$$\omega_j^{pt} = 0 \quad \text{for all pairs } (p, r) \text{ not in } L_j^t. \quad (35)$$

To better understand this fixed-priority recourse structure, first notice that $\sigma_{j0}^{pt} > 0$ indicates that there are some physical containers of type p available for substitutions or export demands. Also remark that

$$\sigma_{j(q-1)}^{pt} \geq 0 \implies \sigma_{jq}^{pt} \geq 0 \quad q = 1, \dots, |L_j^t|.$$

Conversely, $\sigma_{j0}^{pt} < 0$ indicates a deficit of equivalent containers, and

$$\sigma_{j(q-1)}^{pt} \leq 0 \implies \sigma_{jq}^{pt} \leq 0 \quad q = 1, \dots, |L_j^t|.$$

It follows that, through the application of the previous rules, for any given type p , we either have a stock of available "physical" containers of that type (which may eventually become 0), or a deficit of "equivalent" containers of the same type (which may also become 0). Furthermore, substitution of containers of type p^* for type r^* only takes place if containers of type p^* are available ($\sigma_{j(q-1)}^{p^*t} > 0$) and there is a deficit of equivalent containers of type r^* ($\sigma_{j(q-1)}^{r^*t} < 0$). The level of substitutions is bounded by the availability of type p^* , the number of containers required to cover the deficit of type r^* , and the pre-defined upper bound $B_j^{p^*r^*t}$.

To keep on with the tradition of considering non-negative flows, we replace the unrestricted $\tilde{w}_j^{pt}$ variables with pairs of variables $\tilde{w}_{j+}^{pt} \geq 0$ and $\tilde{w}_{j-}^{pt} \geq 0$, representing respectively their positive and negative parts. We are then in a position to completely describe the constraint set of the model. It can be broken down into two parts: the constraint set for the deterministic variables, and the fixed-priority recourse structure which we have just defined: equations (23) (with $\tilde{w}_j^{pt}$ replaced by $(\tilde{w}_{j+}^{pt} - \tilde{w}_{j-}^{pt})$) to (35).

The deterministic constraint set is then made up of equation (22), with $\tilde{w}_j^{pt}$ replaced by $(\tilde{w}_{j+}^{pt} - \tilde{w}_{j-}^{pt})$, of the usual non negativity restrictions on the variables, and

$$\sum_{i' \leq i} \sum_{j \in J_i^{pt} | \tau_j = i - i'} v_{ji}^{pt} = X_i^{pt} - K_i^{pt} \quad \text{for all } i \in I^{pt}, p \in P, t = 1, 2, \dots, T; \quad (36)$$

$$\sum_{j \in J_s^p} v_{sj}^{pt} = Y_s^{pt} \text{ for all } s \in S^{pt}, p \in P, t = 1, 2, \dots, T; \quad (37)$$

$$w_j^{p1} = w_j^{p0} + b_j^{p1} + \sum_{k \in D \cup H \mid \tau_{kj} = 0} w_{kj}^{p1} + K_j^{p1} - \sum_{k \in D \cup H} w_{jk}^{p1} - v_j^{p1} - \sum_{r \in P} w_j^{pr1} \quad (38)$$

for all $j \in D, p \in P$;

$$w_j^{p1} = w_j^{p0} + b_j^{p1} + \sum_{k \in D \cup H \mid \tau_{kj} = 0} w_{kj}^{p1} + K_j^{p1} - \sum_{k \in D \cup H} w_{jk}^{p1} - v_j^{p1} - X_j^{p1} + Y_j^{p1} + e_j^{p1} \quad (39)$$

for all $j \in H, p \in P$;

$$v_j^{p1} + \sum_{r \in P} a_{rp} w_j^{rp1} - \sum_{i \in I_j^{p1}} v_{ji}^1 = 0 \quad \text{for all } j \in D \cup H, p \in P; \quad (40)$$

$$\delta_{jk}^{pt} l_{jk}^{pt} \leq w_{jk}^{pt} \leq \delta_{jk}^{pt} u_{jk}^{pt} \text{ for all } j \in D \cup H, k \in D \cup H, p \in P, t = 1, 2, \dots, T; \quad (41)$$

$$\sum_{i \in \Delta} \delta_{jk}^{pt} = 1 \text{ for all } j \in D \cup H, k \in D \cup H, p \in P. \quad (42)$$

We may now write down the objective of the model, that is to minimize the total expected cost of the empty container distribution system, including substitution costs. It should first be noted, however, that the introduction of the $\bar{w}$ and $\bar{v}$ variables makes the expected recourse independent of all other deterministic flows, except for the stocks at the end of period 1 (w_j^{p1}). It should also be pointed out that recourse actions at a given depot are now independent of flow variables defined for other depots, which makes the expected recourse function separable by depot. The objective can thus be written as:

$$\begin{aligned}
 & \sum_{t=1,2,\dots,T} \sum_{p \in P} \left[\sum_{j \in D \cup H} \left[\sum_{i \in I_j^p} C_{ji}^{pt} (v_{ji}^{pt}) v_{ji}^{pt} + \sum_{k \in D \cup H} C_{jk}^{pt} (\underline{w}_{jk}^t) w_{jk}^{pt} \right] \right. \\
 & \quad \left. + \sum_{s \in S^p} \sum_{j \in J_s^p} C_{sj}^{pt} (v_{sj}^{pt}) v_{sj}^{pt} \right] \\
 & + \sum_{p \in P} \left[\sum_{j \in D \cup H} \left[C_j^{p1} (w_j^{p1}) w_j^{p1} + \bar{C}_j^{p1} (b_j^{p1}) b_j^{p1} + \sum_{r \in P} C_j^{pr1} (w_j^{pr1}) w_j^{pr1} \right] \right. \\
 & \quad \left. + \sum_{j \in H} \bar{C}_j^{p1} (e_j^{p1}) e_j^{p1} \right] \\
 & + \sum_{j \in D} E_{\chi_j, \psi_j} \left[\sum_{t=2,\dots,T} \sum_{p \in P} \left[C_j^{pt} (\underline{\omega}_j^p) \omega_j^{pt} + \bar{C}_j^{pt} (\beta_j^{pt}) \beta_j^{pt} + \sum_{r \in P} C_j^{prt} (\omega_j^{prt}) \omega_j^{prt} \right] \right. \\
 & \quad \left. | \underline{w}_j^1, \underline{\tilde{w}}_j, \underline{\tilde{v}}_j \right] \\
 & + \sum_{j \in H} E_{\chi_j, \psi_j, \xi_j, \phi_j} \left[\sum_{t=2,\dots,T} \sum_{p \in P} \left[C_j^{pt} (\underline{\omega}_j^p) \omega_j^{pt} + \bar{C}_j^{pt} (\beta_j^{pt}) \beta_j^{pt} + \bar{C}_j^{pt} (\epsilon_j^{pt}) \epsilon_j^{pt} + \sum_{r \in P} C_j^{prt} (\omega_j^{prt}) \omega_j^{prt} \right] \right. \\
 & \quad \left. | \underline{w}_j^1, \underline{\tilde{w}}_j, \underline{\tilde{v}}_j \right]
 \end{aligned} \tag{43}$$

5.3 An Equivalent Network Representation

The optimization over the deterministic decision variables can easily be represented in a network format similar to that of Section 5.1, with two nodes for each depot, in each product layer, and for all time periods. For the periods 2 to T , we connect these nodes to a *super-node* (node 0) to allow for the computation of the $\tilde{w}$ and $\tilde{v}$ variables, as illustrated in Figure 6. The expected recourse for each depot is then a function of the flows on the holding arcs between periods 1 and 2 (all other stocks are recourse variables and the corresponding holding arcs should be dropped from the network) and of the flows incident to the super-node.

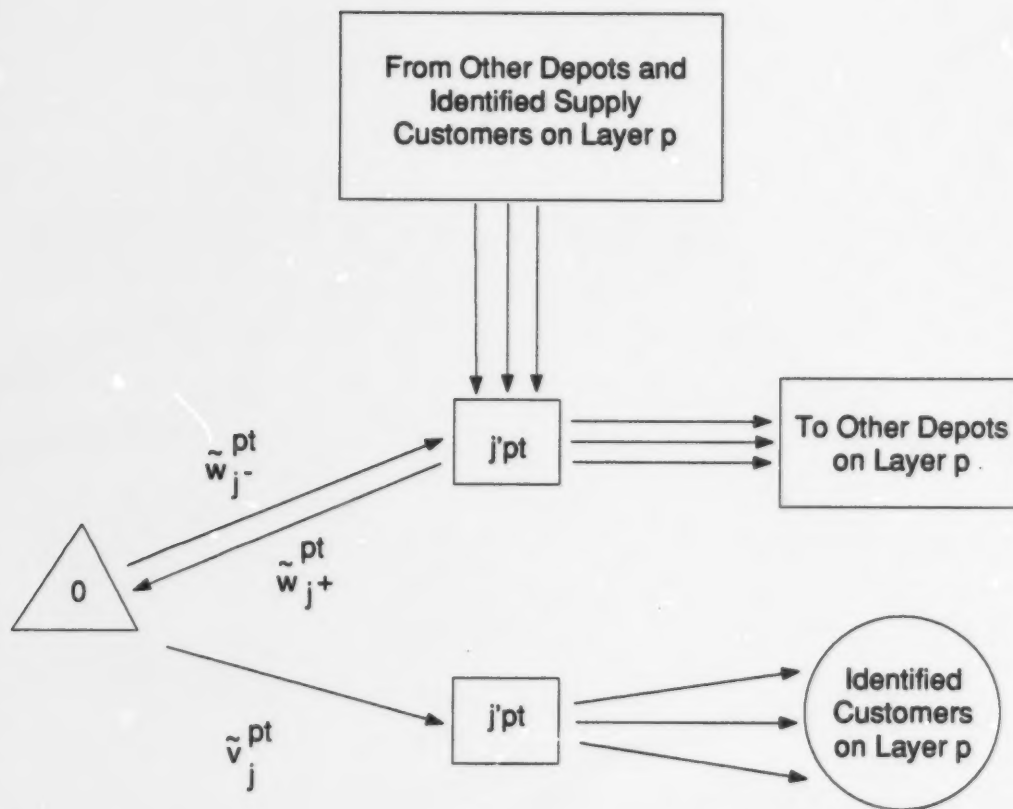


Figure 6. Outline of the Network Representation -
Stochastic Multicommodity Case

6. COMPARISON WITH OTHER FLEET MANAGEMENT PROBLEMS

The empty container allocation problem has received up to now very little attention in the literature, and, to our knowledge, there does not exist any model for that specific problem to which the models described in Sections 3, 4 and 5 could be meaningfully compared. On the other hand, operational fleet management problems arising in other contexts (railroad freight transportation and truckload trucking) have been more extensively studied and a number of models have been developed for them (see Dejax and Crainic, 1987; and Powell, 1988; for detailed reviews on this topic). The purpose of this section is to examine and discuss the similarities and differences between our model and some of the stochastic models which have been proposed in those related settings.

The *stochastic empty railcar distribution* problem is probably the one which most closely resembles the empty container allocation problem. Apart from the fact that it is usually formulated for a single type of railcars at a time (it is thus a single-commodity problem), one has the same general structure of vehicles (railcars in this instance) becoming free which must be distributed to locations (yards) where there is a demand for empties. Among the models that were proposed for this problem, let us mention those of Mendiratta and Turnquist (1982), Jordan and Turnquist (1983), and Beaujon and Turnquist (1991).

The *dynamic vehicle allocation (DVA)* problem has been defined and studied by Powell, Sheffi and Thiriez (1984), Powell (1986, 1987, 1988), Frantzeskakis and Powell (1990), and Powell and Frantzeskakis (1991), in the context of truckload trucking. In this problem, one has to allocate available trucks to potential loads over a time-space network. Its general structure is thus similar to that of the container allocation and the railcar distribution problems, but several elements (most notably, the very high level of uncertainty regarding future demands and the range of options open to the decision-maker) give to the DVA a flavor of its own.

Let us now examine the main aspects on which our container allocation model differs from the aforementioned ones.

Multi-commodity vs single-commodity

The existence of several container types and of complex substitution rules is certainly one of the factors which most distinguishes the container allocation problem from stochastic

dynamic vehicle allocation problems for other transportation modes. As a consequence, the notion of *equivalent container* had to be introduced into the formulation. Furthermore, we had to define a new recourse strategy to allow substitutions in the face of uncertain demands.

Empty only vs combined empty and loaded movements

With regards to the allocation process itself, only empty movements appear as decision variables in the container allocation model (however, as indicated in Section 1, loaded movements are accounted for in the overall planning system and in the routing model), while in many of the previous models (Powell, 1986, 1987, 1988; Beaujon and Turnquist, 1991) both empty and loaded movements are considered. This stems from the fact that loaded containers are shipped outside of the land network (they effectively "disappear" for all planning purposes), while in the other cases the loaded vehicles remain inside the boundaries of the network to become available again for allocation within a short span of time. The same vehicle may thus have to be allocated several times within the planning horizon, which is practically never the case in the land container distribution context. An unexpected consequence of this fact is that current allocation decisions have a lesser impact on the future state of the system under consideration. It is hoped that this may help reduce the feedback distortions caused by the limited planning horizon (see Section 7.3), thus allowing us to work with a shorter planning horizon.

Demand handling

In the container allocation model, all customer demand must absolutely be met within specified delivery windows. In the other models, demands occur at specific instants in time. Requests which cannot be immediately satisfied are either lost (Mendiratta and Turnquist, 1982; Powell, 1986, 1987, 1988) or backordered with a unit penalty per period of delay (Jordan and Turnquist, 1983; Beaujon and Turnquist, 1991). Note also that, in some of the DVA models, it is possible to turn down globally unprofitable loads, even if trucks are available to move them. Consequently, the empty container allocation model is stated as a cost minimization problem, while all models which allow for demand to be lost or turned down must be formulated as profit maximization problems.

It is also important to remark that, in models for other transportation modes, the exact location of demand requests is not specified: these occur at railyards or in cities which are

permanent nodes in the network structure. Hence, the use of specific location of identified customer requests, and of a variable network structure within the allocation process itself, is new.

Particular Characteristics

Under this label, we include the massive balancing flows, the external demands and supplies of empties, the relationships with partner shipping companies, which obviously are all specific features of the container allocation problem, not to be found in any of the other problems.

In summary, it can be said that while the empty container allocation problem presents a general structure which is similar to that of other operational fleet management problems, the models described in this paper differ substantially from previous models proposed for these problems.

7. IMPLEMENTATION ISSUES

Although the main objective of this paper is the presentation of the formulation of the empty container allocation model, several issues related to its implementation and use in actual practice are relevant and are discussed below. We do not aim to be exhaustive. We think, however, that presenting and discussing these items will enhance the understanding of the model and will put our research effort into its proper perspective.

7.1 Modeling context

Several extensions may be envisioned to our modeling framework. Some of these are rather obvious, even though of significant interest for actual applications. Minimum inventory levels at depots and contractual obligations to send empty containers from customer sites back to the original port, are two examples of this case. Some others, such as non linearities in the objective function, are more theoretically interesting, even though in the current state of the available data no practical application may yet be contemplated.

Our interest in this section does not lie, however, with elaborating on this type of model extensions. We are rather concerned with the organizational environment where our model may be applied and the variants thereof. It is clear that the model in its present form corresponds to a centralized organization, where decisions concerning the allocation of empty containers over the entire territory are taken by a unique authority which considers the network as a whole. This is indeed a widespread organizational scheme. Alternate types of organization are, however, equally possible.

One such alternative corresponds to a "bilevel" structure, made up of a central authority and of several regional ones. Here, briefly, the territory is divided into several regions, each of which is managed by its own authority. The latter controls the operations inside its region, as well as the movements with origin inside the region but with destinations outside its borders. The central authority performs the strategic/tactical planing and coordinates the interregional exchanges, especially the massive equilibration flows and the relationships with the outside world.

The model we have presented offers an adequate representation for this case. The variant for a regional authority is a straightforward application of the model, where balancing

movements are treated somewhat similarly to port imports and exports. The central authority model is principally characterized by its representation of the network: the entry and exit points of the system are emphasized, as well as the interdepot balancing arcs, while the regional subnetworks are suitably aggregated. Particular care has to be taken concerning the proper representation of supplies and demands at the nodes of this network, but the present allocation model extends without much effort to the "bilevel" case.

Other types of organization may be imagined and may present a research interest. For example, one may contemplate the case where each regional authority operates nearly autonomously, while the central authority is only concerned with the relationships with the external world. A possible modeling avenue would be to treat the demand/supply balancing problem as an n -person cooperative game. Yet, as we are not aware at the present time of any practical applications of such a model, we do not plan to further explore this alternative in the near future.

7.2 Data requirements

The empty container allocation model, similarly to any other operational model of a large and complex organization, requires for its proper use in an application environment that abundant data be readily available. This information is varied in nature and may be either used directly (e.g., demand figures) or indirectly (e.g., previous routing decisions and their cost implications) to infer correct approximate figures and formulas used by the model. Thus the model has been designed as part of a decision support system for the short term land transportation planning of loaded and empty containers (see Figure 1), and its appropriate implementation assumes the existence of a reliable information system.

The required data falls into one of the following main categories:

- distribution network: the permanent structure (ports, depots), the "active" customers and their characteristics, the corresponding arcs, etc.;
- container types: technical characteristics, substitution rules, availability, limits and costs at partner companies, repairing delays, etc.;
- strategic/tactical guidelines: the permanent network structure, the general allocation of customers to depots, the massive equilibration arcs, etc.;

- container demand and supply: supply and demand forecast models and distribution update procedures, supplies at ports and depots, ships schedules, etc.;
- costs: unit figures and costing formulas for the various activities related to the allocation and distribution of containers.

To implement and run the model requires an appropriate determination of the parameters in all these categories for each period of the rolling horizon. Furthermore, it also requires a careful updating of the data every time the horizon is advanced. To discuss the definition, origin, evaluation/approximation and update procedures for all data items required by the allocation model is clearly beyond the scope of this paper. Besides, some of these points have already been mentioned in previous sections, while some others are rather straightforward. We want, however, to address some of the issues related to specifying demands, supplies and costs, due to their importance to the numerical tractability of the model and its actual utilization.

Accurate and constantly updated forecasts of the supply and demand of empty containers are critical for a successful implementation of the allocation model. Preliminary studies (Gendreau et al., 1990) have shown that simple forecasting methods cannot capture the complexity of the demand mechanism at work and similar conclusions could probably be drawn with regards to the supplies. A more careful analysis reveals that observed demands and supplies are indeed the result of the superposition of various processes with widely different characteristics. To obtain good forecasts, it is thus necessary to identify as clearly as possible these processes, to monitor them closely through the management information system of the company, to derive good individual forecasts by process, and finally to merge all these forecasts into specific probability distributions. Finally, these distributions must be updated on an individual basis from day-to-day.

With respect to costs, relevant questions are: "what unit costs to use?", "what functional form to specify", "are these costs separable by commodity layer?". We briefly examine each of these issues for the main cost categories of the model. First, note that *handling* costs, corresponding to the various operations and intermodal transshipment of containers at ports, depots and customer sites do not appear in our formulations. Indeed, these costs are well determined and make up for a significant part of the total cost of the system. Yet, for the most part, they may not be avoided, and, furthermore, their optimization is more a

strategic/tactical issue rather than an operational one. Thus, even if these costs have to be accounted for in order to get a realistic picture of the company's performance, they are not part of the allocation decision process.

Unit *holding* costs for the storage of containers at depots are generally well known; these may be specific to given depot and container type, but are usually invariant for the duration of the planning horizon. Hence, the total holding cost is usually linear and separable. The same is also true for *substitution* costs between different container types, independently of their representation of actual figures or "penalty" charges if certain substitutions are possible but not desirable.

Transportation costs represent the most significant part of the total system costs and, thus, of the objective function of our model. Yet, as already mentioned, maritime container shipping companies do not usually own land transportation means. They therefore have to contract the services of carriers of various modes. In the case of trucking, they use the services of a large number of local companies, and the cost of transportation is often determined on a case by case basis, according to many factors such as the geographical area, the technical association of truck size and type and container types, the distance to travel, the insurance of return freight or paid return, etc. For rail shipments, they call for the services of national railway companies, which propose an equally complex tariff structure, where the cost of transportation depends on the origin and destination zones, the type, size, weight and number of containers to be shipped together, etc.

We have initiated an investigation aimed at determining formulae for modal unit transportation costs, which take into account the particularities mentioned in the previous paragraph and which are based on historical data. The initial results are extremely encouraging (Vasseur, 1989; Crainic, Dejax and Gendreau, 1989). These formulae may be used to estimate the unit costs to be included in the allocation model. However, for any given origin-destination pair, the particular transportation arrangement may vary from period to period, especially when trucking is involved. In particular, two containers may be moved by the same truck, or the same vehicle may be used in a cyclic route to move several distinct containers. However, this routing information is not available at allocation time. Thus, we plan to integrate a cost update mechanism to the decision support system, to re-estimate the unit costs prior to the advancement of the planning horizon, based on the actual routing decisions.

For the allocation model, we propose to use linear transportation cost functions based on the previously defined unit costs. This is a reasonable hypothesis for an operational model, since most cost nonlinearities, such as potential economies of scale offered by certain transportation modes, have been considered by the strategic/tactical model. This hypothesis also applies when containers of various types are transported together (e.g., by rail or barge), in particular the case of the massive balancing flows and traffic among ports and depots. Hence, we may assign to each arc, in each particular commodity layer, its respective function which depends only on the particular flow on the arc. Therefore, function separability by link is insured.

7.3 Ending the planning horizon

The models of the previous sections cover only a planning horizon of T periods, while the operations of the company will extend in the future for a much longer span of time. This situation, which is encountered in almost all dynamic models, requires careful scrutiny since one would like to minimize (hopefully nullify) the distortions created by the use of a finite horizon model to represent an, for all practical purposes, infinite horizon reality.

Obviously, using the model in a rolling horizon mode helps to mitigate the negative effects of this situation, but it does not by itself eliminate them. It is well-known that, for certain types of models, end-of-horizon distortions will tend to feed back on the values of the decision variables, even in the first period (see, for instance, Hughes and Powell 1988, for a more detailed discussion on this topic). It is thus very important to carefully select the length of the planning horizon and the *end-of-horizon conditions*, in order to obtain valid solutions from the model.

With regards to the length of the planning horizon, two key factors have to be taken into account. One must first be aware that, for any real-life application of the model, the number of decision variables in any period will already be fairly large. This implies that the length of the planning horizon must be limited to some reasonable value (say between 10 and 20 days), if the model is to be computationally tractable.

The second factor to be considered is the information available on the future supply and demand of empty containers. As we go down the time axis, we move from an area where a

large portion of the supply and demand requests can be deemed as certain, to one where everything (or almost) is uncertain. Ideally, the length of the planning horizon should be such as to cover certain demands and supplies (identified customer requests and known import and export movements).

If such an ideal planning horizon is possible, one may then specify fairly simple end-of-horizon conditions to prevent unrealistic solutions. These conditions should be such as to force reasonable values for the empty container stocks in each depot at the end of period T . (These values may be derived from historical experience and from practical operational constraints, or from an empty container inventory model.) In the deterministic version of the model, this can be accomplished by imposing lower and, eventually, upper bounds on these stocks:

$$l_j^{pT} \leq w_j^T \leq u_j^{pT} \quad \text{for all } j \in D \cup H, p \in P. \quad (44)$$

One could also include a *salvage value* term in the holding cost functions for period T ($C_j^T(w_j^T)$) to account for the marginal value of the various container types at each depot (Hughes and Powell, 1988). This last option could also be implemented in the stochastic formulation. Obviously, in any practical application, some experimentation with possible values for these bounds and salvage terms would be necessary to ascertain that the chosen values are indeed correct.

When no feasible ideal horizon exists, the specification of the end-of-horizon conditions becomes a much more difficult task. In such a situation, one would first have to deal with "split delivery windows", i.e., delivery windows which fall partly within the planning horizon and partly outside of it. Deciding whether or not the corresponding demands should be included in the planning horizon is not obvious. In most cases, one would probably be well-advised to exclude them and to slightly adjust the lower bounds on the empty container stocks at the depots from which these demands should be satisfied later on (this is known from the strategic/tactical plan). Similar remarks apply to known import supplies and export demands occurring after the last period of the horizon.

In summary, it must be stressed that there is no cut and dry answer to the problems created by the necessity to work with a limited planning horizon. The model offers a number of means (bounds, salvage values) which can be used to tackle these problems, but the

specific solution to implement in any practical application of the model is highly dependent upon the situation at hand.

7.4 Algorithmic perspectives

Even though the development and appraisal of algorithms for the proposed allocation model are clearly beyond the scope of this paper, we would now like to briefly examine and discuss possible solution approaches.

In Section 5.3, we indicated how the stochastic multi-commodity model could be represented as a network flow problem with respect to the deterministic decision variables. Under the assumptions made earlier (i.e. exogeneous determination of the massive equilibration flow control variables δ -- see Section 3.3. -- and linear cost functions -- see Section 7.2), the costs on most links of this network are link-separable and linear. In fact, the only real difficulty in solving this problem comes from the fact that the expected recourse function depends jointly on the stocks at the end of the first period and on the flows incident to the super-node. However, with linear holding, borrowing, unmet demand and substitution costs, this function will be piecewise linear with respect to any of these variables, when all others are fixed. Furthermore, it should be convex if the cost functions and the priority structures are economically sensible actions in a given setting (an example of inconsistency would be having a rule prescribing substitutions of containers of type p for type r when it would be cheaper to borrow from partner companies). The evaluation of derivatives (or one-sided derivatives at break points) is obviously a tedious task, since it requires in general the assessment of all future impacts of modifying a variable (the periods are linked by the inventory recourse variables), but, given the depot-separability of the expected recourse function, we believe that it can be achieved at reasonable computational effort if clever programming techniques are used. It would then become possible to solve the problem by customizing standard descent algorithms such as the Frank-Wolfe method.

In a practical setting, an attractive alternate approach would consist in solving the problem as a deterministic one, replacing the random supplies and demands by their expected values. In their comparison of deterministic and stochastic formulations for the Dynamic Vehicle Allocation (DVA) problem, Frantzeskakis and Powell (1990) have concluded that stochastic models were clearly superior to deterministic ones. It is not obvious whether such

conclusions could be extended to the empty container allocation problem. For one thing, as was explained in Section 2.6, demand is not handled in the same fashion in both problems: in the DVA, less profitable loads may (and indeed should) be turned down, while in our model all demand must be satisfied. This may very well mitigate the superiority of the stochastic formulations.

Should one decide to solve the problem as a deterministic one, a number of issues still remain to be addressed. Several of these were discussed by Frantzeskakis and Powell in the context of the DVA: fractional optimal solutions, even for the first period flows, which cannot be implemented in practice; selection of proper planning horizon length (since end effects distortions are amplified in a deterministic setting); choice of appropriate salvage values; ... The answers they have provided to these questions remain valid for our problem and will not be discussed any further.

There is however one issue which is specific to the empty container allocation problem and which deserves careful consideration: the choice of an efficient solution approach for the deterministic version of the model, since this is still a large dynamic multicommodity network flow with gains problem, which is not easily solved. Obviously, one could apply a standard network flow with gains algorithm to the multilayer representation of the network, thus solving the problem as a single-commodity one. However, this approach does not account for the very special structure of the network. More precisely, it does not acknowledge the fact that, were it not for the substitution arcs, the problem would decompose naturally in a series of much smaller pure network problems. In this context, decomposition methods which seek to exploit this near-separability of the problem seem quite attractive. Among these, *variable splitting/Lagrangian decomposition* schemes la Guignard-Spielberg (1988), in which each substitution flow would be represented by two variables with the linking constraint being dualized, appear particularly promising.

To close this section, it should be pointed out that the evaluation of algorithms for the proposed model should be conducted using a *rolling horizon simulation procedure* similar to the one devised by Frantzeskakis and Powell, since the model was explicitly developed for a rolling horizon decision setting.

CONCLUSIONS

We have described a dynamic, stochastic model for empty container allocation. Although based on actual experience with a large container shipping company, it offers a general formulation for this class of problems. The model reflects the operational and planning complexity of the land component of the international maritime shipping of containers, particularly the space and time dependency of events and the uncertainty of supply and demand data. In particular, it addresses the operational characteristics which are specific to empty container allocation, such as substitutions, relationships with partner companies, imports and exports, massive equilibration flows, etc.

To our knowledge, this is the first model specifically aimed at the empty container land allocation problem. We have attempted, therefore, to explore as comprehensively as possible the various modeling issues related to this problem, while proposing stochastic formulations that are computationally tractable. This has allowed us to highlight the originality of the problem and to explain the differences which may be observed between our modeling framework and formulations for similar allocation problems arising in the context of other transportation modes.

We have sketched some of the most promising avenues for the development of an efficient solution technique for the proposed model. This, together with a continuing exploration of richer stochastic formulations, constitutes the next stage of our research program. We also plan to start the investigation of the loaded and empty container routing problem and that of the operational planning system required for an efficient actual implementation of the models.

ACKNOWLEDGEMENTS

The authors wish to thank the Area Editor, the Associate Editor and the referees for their thorough review of the manuscript; their suggestions and comments have helped us write a better paper. The authors are also grateful to the Canadian *National Science and Engineering Research Council*, to the *Fonds F.C.A.R.* of the Province of Québec, and to the *Programme de coopération France - Québec*, who have provided the financial support for this research.

REFERENCES

- Antonisse K. 1988, A Decision Support System for Container Replenishment, TIMS XXVIII/EURO IX Joint Meeting, Paris, July 6-8.
- Antush R.M. 1989, Research Issues at Sealand for the 1990's, ORSA/TIMS Joint National Meeting, New York, October 16-18.
- Beaujon G.J. and Turnquist M.A. 1991, A Model for Fleet Sizing and Vehicle Allocation, *Transportation Science* 25(1), 19-45.
- Breton M. and El Hachem S. 1991, Decomposition Algorithm for Stochastic Dynamic Programs, Report G-91-15, GERAD, Ecole des Hautes Etudes Commerciales, Montréal.
- Crainic T.G., Dejax P.J. and Delorme L. 1989, Models for Multimode Multicommodity Location Problems with Interdepot Balancing Requirements, *Annals of Operations Research* 18, 279-302.
- Crainic T.G., Gendreau M. and Dejax P.J. 1990, Modelling the Container Fleet Management Problem Using a Stochastic Dynamic Approach, *Operational Research'90* (H.E. Bradley Editor), 473-486, Pergamon Press.
- Crainic T.G., Dejax P.J. and Gendreau M. 1989, Transport terrestre multimodal de conteneurs: problématique et modèles, in *Logistique/Logistics: Production, Distribution, Transport / Manufacturing, Distribution, Transportation*, Proceeding of the Joint Conference on the Practice and Theory of Operations Management and 4es journées francophones sur la Logistique et les Transports, 105-113, AFCET, Paris.
- Crainic T.G., Dejax P.J., Gendreau M. and Benamar F. 1991, Multimodal Multiproduct Combined Vehicle Routing - Traffic Itinerary Generation, ORSA/TIMS Joint National Meeting, Anaheim, November 3-7.
- Dejax P.J. and Crainic T.G. 1987, A Review of Empty Flows and Fleet Management Models in Freight Transportation, *Transportation Science* 21(4), 227-247.
- Dejax P.J., Crainic T.G. and Delorme L. 1988, Strategic/Tactical Planning of Container Land Transportation Systems, TIMS XXVIII/EURO IX Joint Meeting, Paris, July 6-8.
- Guignard M. and Spielberg K. 1985, Separable Lagrangean Relaxation for Weakly-Linked Mixed Integer Programming Problems, *Methods of Operations Research* 49, 239-253.
- Hughes R.E. and Powell W.B. 1988, Mitigating End Effects in the Dynamic Vehicle Allocation Model, *Management Science* 34(7), 859-879.
- Frantzeskakis L.F. and Powell W.B. 1989, A Successive Linear Approximation Procedure for Stochastic, Dynamic Vehicle Allocation Problems, *Transportation Science* 24(1), 40-57.
- Gendreau M., Crainic T.G., Dejax P. and Steffan H. 1990, Forecasting Short-Term Demand for Empty Containers: A Case Study, Publication #714, Centre de recherche sur les transports, Université de Montréal, forthcoming *Transportation Research Records*.

- Jarke M. 1982, Developing Decision Support Systems: A Container Management Example, *Int. J. Policy Anal. Inform. Syst.* 6, 351-372.
- Jordan W.C. and Turnquist M.A. 1983, A Stochastic, Dynamic Network Model for Railroad Car Distribution, *Transportation Science* 17, 123-145.
- Mendiratta V.B. and Turnquist M.A. 1982, A Model for the Management of Empty Freight Cars, *Transportation Research Record* 838, 50-55.
- Powell W.B. 1986, A Stochastic Model of the Dynamic Vehicle Allocation Problem, *Transportation Science* 20(2), 117-129.
- Powell W.B. 1987, An Operational Planning Model for the Dynamic Vehicle Allocation Problem with Uncertain Demands, *Transportation Research* 21B, 217-232.
- Powell W.B. 1988, A Comparative Review of Alternative Algorithms for the Dynamic Vehicle Allocation Problem, in: *Vehicle Routing : Methods and Studies* (B.L. Golden and A.A. Assad, Editors), pp. 249-291, North-Holland, Amsterdam.
- Powell W.B., Sheffi Y. and Thiriez S. 1984, The Dynamic Vehicle Allocation Problem with Uncertain Demands, in *9th Int. Symp. on Transportation and Traffic Theory* (J. Volmuller and R. Hamerslag, Editors), VNU Science Press.
- Powell W.B. and Frantzeskakis L.F. 1991, Restricted Recourse Strategies for Dynamic Network with Random Arc Capacities, Report SOR-89-24, School of Engineering and Applied Science, Department of Civil Engineering and Operations Research, Princeton University.
- Vasseur C. 1989, Les coûts du transport ferroviaire par conteneurs en France: étude de cas, Publication # 650, Centre de recherche sur les transports, Université de Montréal.